GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:49:50 ; Search time 505.437 Seconds
(without alignments)
8094.223 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	137	100.0	373	10	AW580404	AW580404 PM2-HT045
c 2	137	100.0	396	14	H89576	H89576 yw28c12.r1
c 3	137	100.0	455	9	AA427831	AA427831 zw49d12.r
c 4	137	100.0	462	14	CB045405	CB045405 NISC_gc10
c 5	137	100.0	473	14	CA389721	CA389721 cs101f03.
c 6	137	100.0	489	28	AQ015185	AQ015185 CIT-HSP-2
c 7	137	100.0	647	10	AW664684	AW664684 hi84e10.x
c 8	137	100.0	696	10	AW950480	AW950480 EST362550
c 9	137	100.0	710	9	AV725574	AV725574 AV725574
c 10	137	100.0	727	12	BG567176	BG567176 602589640
c 11	137	100.0	740	14	CD356811	CD356811 AGENCOURT
c 12	137	100.0	745	14	CD357556	CD357556 AGENCOURT
c 13	137	100.0	1121	12	BM546012	BM546012 AGENCOURT
c 14	136	99.3	743	14	CD110642	CD110642 AGENCOURT
c 15	135.4	98.8	375	10	AW607064	AW607064 PM2-HT045
c 16	135.4	98.8	862	14	CB962272	CB962272 AGENCOURT
c 17	135.4	98.8	911	12	BI752714	BI752714 603028327
18	120	87.6	352	10	AW385154	AW385154 PM2-HT045
19	111.4	81.3	353	10	AW580423	AW580423 PM2-HT045
c 20	107	78.1	631	14	CB465228	CB465228 726497 MA
c 21	100.6	73.4	291	10	BB339107	BB339107 BB339107
22	100.6	73.4	390	9	AI324317	AI324317 mq87c07.y
c 23	100.6	73.4	607	10	BE370292	BE370292 601222549
24	100.6	73.4	610	9	AA146038	AA146038 mq87c07.r
c 25	100.6	73.4	735	14	CB570707	CB570707 AGENCOURT
c 26	100.6	73.4	4437	11	AK036936	AK036936 Mus muscu
c 27	99	72.3	296	10	BB041920	BB041920 BB041920
c 28	98.6	72.0	643	14	CB850037	CB850037 MRA-1794
c 29	97.4	71.1	313	10	BB332959	BB332959 BB332959
c 30	95.8	69.9	210	10	BB171938	BB171938 BB171938
c 31	95.8	69.9	268	10	BB178429	BB178429 BB178429
c 32	95.8	69.9	291	10	BB174122	BB174122 BB174122
c 33	95.8	69.9	329	10	BB085328	BB085328 BB085328
c 34	94.6	69.1	841	10	BF791019	BF791019 602251072
c 35	94.2	68.8	293	9	AV307178	AV307178 AV307178
c 36	92.6	67.6	306	10	BB086855	BB086855 BB086855
c 37	92.6	67.6	332	10	BB235445	BB235445 BB235445
c 38	89.6	65.4	418	13	BY377466	BY377466 BY377466
c 39	86.2	62.9	239	9	AV232779	AV232779 AV232779
c 40	86.2	62.9	293	10	BB307434	BB307434 BB307434
c 41	80.8	59.0	641	9	AI528613	AI528613 mq87c07.x
c 42	78.4	57.2	945	10	BF180441	BF180441 601805231
43	72	52.6	955	13	BX419179	BX419179 BX419179
c 44	68.4	49.9	651	13	BU338697	BU338697 603515030
c 45	68.4	49.9	733	13	BU257193	BU257193 603744569

ALIGNMENTS

RESULT 1
AW580404/c
LOCUS AW580404 373 bp mRNA linear EST 16-MAR-2000
DEFINITION PM2-HT0451-080100-003-h09 HT0451 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW580404
VERSION AW580404.1 GI:7255453
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 373)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM2&t2=PM2-HT0451-080100-003-h09&t3=2000-01-08&t4=1)

Seq primer: puc 18 forward
High quality sequence stop: 373.

FEATURES Location/Qualifiers
source 1. .373
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0451"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 100.0%; Score 137; DB 10; Length 373;
Best Local Similarity 100.0%; Pred. No. 4.1e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 310

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 250

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        249 TTTGATGTTAATTAATT 233
```

RESULT 2
H89576/c
LOCUS H89576 396 bp mRNA linear EST 28-NOV-1995
DEFINITION yw28c12.r1 Morton Fetal Cochlea Homo sapiens cDNA clone IMAGE:253558 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.
ACCESSION H89576
VERSION H89576.1 GI:1079922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 300
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1088 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 300.

FEATURES Location/Qualifiers
source 1. .396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3891011"
/db_xref="taxon:9606"
/clone="IMAGE:253558"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned unidirectionally. Primer: Oligo dT. Fetal cochlea, normal. 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP XR Vector. Library constructed by N. Robertson, C. Morton. ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.1e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 205

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 145

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        144 TTTGATGTTAATTAATT 128
```

RESULT 3
AA427831/c
LOCUS AA427831 455 bp mRNA linear EST 16-OCT-1997
DEFINITION zw49d12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:773399 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.
ACCESSION AA427831
VERSION AA427831.1 GI:2111628
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 391.

FEATURES Location/Qualifiers
source 1. .455
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:773399"

/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 100.0%; Score 137; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 4e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 96

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 36

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         35 TTTGATGTTAATTAATT 19
```

RESULT 4
CB045405/c
LOCUS CB045405 462 bp mRNA linear EST 17-JAN-2003
DEFINITION NISC_gc10c01.y1 NCI_CGAP_Co17 Homo sapiens cDNA clone IMAGE:3218281 5', mRNA sequence.
ACCESSION CB045405
VERSION CB045405.1 GI:27783692
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 462)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM7870 row: E column: 2
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1. .462
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3218281"
/tissue_type="juvenile granulosa tumor"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Co17"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 462;
Best Local Similarity 100.0%; Pred. No. 4e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 373

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 313

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        312 TTTGATGTTAATTAATT 296
```

RESULT 5
CA389721/c
LOCUS CA389721 473 bp mRNA linear EST 06-NOV-2002
DEFINITION cs101f03.y3 Human Retinal pigment epithelium/choroid cDNA (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs101f03 5', mRNA sequence.
ACCESSION CA389721
VERSION CA389721.1 GI:24720152
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A., Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: Over 6000 non-redundant transcripts, novel genes and splice variants
JOURNAL Mol. Vis. 8 (4), 205-220 (2002)
MEDLINE 22103460
PUBMED 12107410
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 101 row: f column: 03
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1. .473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="cs101f03"
/tissue_type="RPE/choroid"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA (Un-normalized, unamplified): cs"
/note="Organ: Eye; Vector: pCMVSPORT6; Two different donor eyes (75-80 years old) yielded approximately 600 mg of dissected RPE/choroid tissue. This in turn yielded 340 ug of total RNA and 7 ug of mRNA. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies (Rockville, MD; now part of Invitrogen Corp), essentially following the protocols of the SuperScript Plasmid System (Invitrogen Corp. <http://www.invitrogen.com/>). The library code designation was cs. For this library, cDNA inserts were cloned into the NotI/MluI sites of the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

```
ORIGIN

  Query Match            100.0%;  Score 137;  DB 14;  Length 473;
  Best Local Similarity  100.0%;  Pred. No. 3.9e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 172

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 112

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        111 TTTGATGTTAATTAATT 95


RESULT 6
AQ015185/c
LOCUS       AQ015185                489 bp    DNA     linear   GSS 09-JUN-1998
DEFINITION  CIT-HSP-2310F4.TR CIT-HSP Homo sapiens genomic clone 2310F4,
            genomic survey sequence.
ACCESSION   AQ015185
VERSION     AQ015185.1  GI:3193921
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2310F4.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2310F4"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match            100.0%;  Score 137;  DB 28;  Length 489;
  Best Local Similarity  100.0%;  Pred. No. 3.9e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 299

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 239

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        238 TTTGATGTTAATTAATT 222


RESULT 7
AW664684/c
LOCUS       AW664684                647 bp    mRNA    linear   EST 06-APR-2000
DEFINITION  hi84e10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2979018 3' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA
            sequence.
ACCESSION   AW664684
VERSION     AW664684.1  GI:7457227
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2979018"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match            100.0%;  Score 137;  DB 10;  Length 647;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 92

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 32

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         31 TTTGATGTTAATTAATT 15


RESULT 8
AW950480/c
```

```
LOCUS       AW950480                696 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST362550 MAGE resequences, MAGA Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW950480
VERSION     AW950480.1  GI:8140134
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 15
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGA"
                     /note="Vector: pBluescriptSKm"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 10;  Length 696;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 513

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 453

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        452 TTTGATGTTAATTAATT 436


RESULT 9
AV725574/c
LOCUS       AV725574                710 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725574 HTC Homo sapiens cDNA clone HTCBTD10 5', mRNA sequence.
ACCESSION   AV725574
VERSION     AV725574.1  GI:10831099
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HTCBTD10"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="HTC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 9;  Length 710;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 337

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 277

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        276 TTTGATGTTAATTAATT 260


RESULT 10
BG567176/c
LOCUS       BG567176                727 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602589640F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723833 5',
            mRNA sequence.
ACCESSION   BG567176
VERSION     BG567176.1  GI:13574829
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1583  row: p  column: 10
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4723833"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
```

5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

```
  Query Match             100.0%;  Score 137;  DB 12;  Length 727;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 278

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 218

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        217 TTTGATGTTAATTAATT 201
```

RESULT 11
CD356811/c
LOCUS CD356811 740 bp mRNA linear EST 29-MAY-2003
DEFINITION AGENCOURT_14253503 NIH_MGC_187 Homo sapiens cDNA clone IMAGE:30404795 5', mRNA sequence.
ACCESSION CD356811
VERSION CD356811.1 GI:31128222
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 740)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM190 row: o column: 12
High quality sequence stop: 578.
FEATURES Location/Qualifiers
source 1. .740
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30404795"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_187"
/note="Organ: Blood vessels - aorta, basilar and artery; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

```
  Query Match             100.0%;  Score 137;  DB 14;  Length 740;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 491

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 431

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        430 TTTGATGTTAATTAATT 414
```

RESULT 12
CD357556/c
LOCUS CD357556 745 bp mRNA linear EST 29-MAY-2003
DEFINITION AGENCOURT_14253482 NIH_MGC_187 Homo sapiens cDNA clone IMAGE:30402369 5', mRNA sequence.
ACCESSION CD357556
VERSION CD357556.1 GI:31128967
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM184 row: j column: 10
High quality sequence stop: 460.
FEATURES Location/Qualifiers
source 1. .745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30402369"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_187"
/note="Organ: Blood vessels - aorta, basilar and artery; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

```
  Query Match             100.0%;  Score 137;  DB 14;  Length 745;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-25;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 201

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 141

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        140 TTTGATGTTAATTAATT 124


RESULT 13
BM546012/c
LOCUS       BM546012                1121 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6505286 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5588419
            5', mRNA sequence.
ACCESSION   BM546012
VERSION     BM546012.1  GI:18778623
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12359  row: h  column: 20
            High quality sequence start: 50
            High quality sequence stop: 764.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5588419"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
ORIGIN

  Query Match             100.0%;  Score 137;  DB 12;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 455

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 395

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        394 TTTGATGTTAATTAATT 378


RESULT 14
CD110642/c
LOCUS       CD110642                743 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  AGENCOURT_13995397 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30373580 5', mRNA sequence.
ACCESSION   CD110642
VERSION     CD110642.1  GI:30754851
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM167  row: j  column: 21
            High quality sequence stop: 578.
FEATURES             Location/Qualifiers
     source          1. .743
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30373580"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: Blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.5-4.0 kb). 14/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             99.3%;  Score 136;  DB 14;  Length 743;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-25;
  Matches  136;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 79

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 19

Qy        121 TTTGATGTTAATTAAT 136
              ||||||||||||||||
Db         18 TTTGATGTTAATTAAT 3


RESULT 15
AW607064/c
```

```
LOCUS       AW607064                375 bp    mRNA    linear   EST 23-MAR-2000
DEFINITION  PM2-HT0451-170100-004-h12 HT0451 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW607064
VERSION     AW607064.1  GI:7311805
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM2&t2=PM2-HT0451-
            170100-004-h12&t3=2000-01-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0451"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             98.8%;  Score 135.4;  DB 10;  Length 375;
  Best Local Similarity   99.3%;  Pred. No. 1.1e-24;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 312

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCACAATTATCTGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 252

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        251 TTTGATGTTAATTAATT 235


Search completed: April  6, 2004, 14:12:02
Job time : 511.437 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 02:56:20 ; Search time 64.812 Seconds
(without alignments)
8979.866 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	137	100.0	137	6	ABK86496	Abk86496 Human Apo
c	2	137	100.0	158	6	ABK86469	Abk86469 Human apo
c	3	137	100.0	200	6	ABK86468	Abk86468 Human dys
c	4	137	100.0	455	9	ADD32504	Add32504 Human mit
	5	137	100.0	996	6	ABK86497	Abk86497 Human Apo
	6	137	100.0	1230	6	ABK86462	Abk86462 Human Apo
	7	137	100.0	1234	6	ABK86463	Abk86463 Human Apo
c	8	137	100.0	2691	6	ABK81996	Abk81996 Human dys
c	9	137	100.0	8689	6	ABK82000	Abk82000 DNA encod
c	10	137	100.0	11443	6	ABK82002	Abk82002 DNA encod
c	11	137	100.0	13957	6	ABK81959	Abk81959 cDNA enco
c	12	137	100.0	13957	6	ABT10904	Abt10904 Human bre
c	13	137	100.0	13957	6	ABN95786	Abn95786 Gene #228
c	14	137	100.0	13957	6	ABS69900	Abs69900 Human dys
c	15	126	92.0	13977	6	ABS70403	Abs70403 Human bon
c	16	107	78.1	959	6	ABN74601	Abn74601 Bovine em
c	17	100.6	73.4	13815	2	AAV18885	Aav18885 Mus muscu
c	18	100.6	73.4	13815	6	ABK81960	Abk81960 cDNA enco
c	19	100.6	73.4	13815	6	ABI99799	Mouse isc
	20	100.6	73.4	19307	2	AAT27558	Aat27558 Shuttle v
	21	57.2	41.8	108	6	ABK86467	Abk86467 Human apo
	22	41	29.9	60	6	ABK86471	Abk86471 Human apo
	23	33.6	24.5	1863	9	ADE60952	Ade60952 Human gen
c	24	31.8	23.2	6428	4	AAK78562	Aak78562 Human imm
c	25	31.8	23.2	96591	9	ADC85301	Adc85301 Mouse Sos
c	26	31.8	23.2	96592	8	ADA02822	Ada02822 Human SOS
c	27	31.8	23.2	96592	9	ADB72560	Adb72560 Human SOS
c	28	31.6	23.1	2000	6	ABZ17261	Abz17261 Arabidops
	29	31.2	22.8	1977	6	ABL62657	Abl62657 Colon ade
	30	31.2	22.8	1977	6	ABT10871	Abt10871 Human bre
	31	31.2	22.8	1977	6	ABK83769	Abk83769 Human cDN
	32	31.2	22.8	1979	2	AAT87015	Aat87015 Human Fc-
	33	31.2	22.8	1990	1	AAN92451	Aan92451 cDNA inse
	34	31.2	22.8	2461	6	AAS94845	Aas94845 Human DNA
c	35	30.8	22.5	47319	4	AAK72230	Aak72230 Human imm
c	36	30.8	22.5	47319	4	AAK64813	Aak64813 Human imm
	37	30.8	22.5	236303	4	AAS11614	Aas11614 Human gen
c	38	30.6	22.3	538	6	ABL78454	Abl78454 Human ova
c	39	30.6	22.3	539	6	ABK62246	Abk62246 Rat seque
c	40	30.6	22.3	539	7	ABT40428	Abt40428 Toxicity
c	41	30.6	22.3	539	9	ADB55325	Adb55325 Toxicity-
c	42	30.6	22.3	539	9	ADB49833	Adb49833 Primary r
	43	30.6	22.3	6644	2	AAX33181	Aax33181 Base sequ
	44	30.6	22.3	7372	2	AAX33182	Aax33182 Base sequ
	45	30.6	22.3	7797	2	AAX33180	Aax33180 Cowpox vi

ALIGNMENTS

RESULT 1
ABK86496
ID ABK86496 standard; DNA; 137 BP.
XX
AC ABK86496;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human Apo-dystrophin-4 inversion sequence.
XX
KW Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3. .137
FT /*tag= a
FT /product= "Apo-dystrophin-4 peptide appearing as
FT AAU98738"
FT /partial
FT /note= "No start or stop codon shown"
FT /transl_except= (pos:21. .23,aa:Xaa)
FT /transl_except= (pos:48. .50,aa:Xaa)
FT /transl_except= (pos:93. .95,aa:Xaa)
FT /transl_except= (pos:123. .125,aa:Xaa)
FT /transl_except= (pos:129. .131,aa:Xaa)
FT /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN GB2368064-A.
XX
PD 24-APR-2002.
XX
PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
DR P-PSDB; AAU98737.
XX
PT New human regulatory polynucleotide, useful for treating disorders

PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Claim 1; Page 169; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is the apo-dystrophin-4 inversion sequence and upstream
CC genomic region
XX
SQ Sequence 137 BP; 56 A; 19 C; 32 G; 30 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.6e-33;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        121 TTTGATGTTAATTAATT 137
```

RESULT 2
ABK86469/c
ID ABK86469 standard; cDNA; 158 BP.
XX
AC ABK86469;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human apo-dystrophin-4 cDNA fragment with inversion breakpoint #2.
XX
KW Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia; dystrophin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_recomb 137
FT /*tag= a
FT /label= Inversion_breakpoint
XX
PN GB2368064-A.
XX
PD 24-APR-2002.
XX
PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
XX
PT New human regulatory polynucleotide, useful for treating disorders
PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Disclosure; Fig 16B; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is a apo-dystrophin-4 cDNA fragment showing an inversion
CC breakpoint (recombination signal sequence) similar to that in with human
CC dystrophin
XX
SQ Sequence 158 BP; 35 A; 35 C; 20 G; 68 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 158;
Best Local Similarity 100.0%; Pred. No. 1.7e-33;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 78

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 18

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         17 TTTGATGTTAATTAATT 1
```

RESULT 3
ABK86468/c
ID ABK86468 standard; DNA; 200 BP.
XX
AC ABK86468;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human dystrophin genomic DNA with inversion breakpoint #2.
XX
KW Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia; dystrophin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_recomb 42

```
FT                   /*tag=  a
FT                   /label= Inversion_breakpoint
FT   misc_recomb     179
FT                   /*tag=  b
FT                   /label= Inversion_breakpoint
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 16B; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is a human dystrophin genomic DNA fragment showing
CC   inversion breakpoints (recombination signal sequence) similar to that in
CC   apo-dystrophin-4
XX
SQ   Sequence 200 BP; 44 A; 40 C; 28 G; 88 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 120

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 60

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         59 TTTGATGTTAATTAATT 43


RESULT 4
ADD32504/c
ID   ADD32504 standard; DNA; 455 BP.
XX
AC   ADD32504;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human mitochondrial DNA sequence SEQ ID NO:274.
XX
KW   ds; human; array; mitochondrial; hybridisation; energy-metabolism;
KW   mitochondrial disease; oxidative phosphorylation dysfunction;
KW   oxidative stress; apoptosis; aging.
XX
OS   Homo sapiens.
XX
PN   WO2003020220-A2.
XX
PD   13-MAR-2003.
XX
PF   30-AUG-2002; 2002WO-US027886.
XX
PR   30-AUG-2001; 2001US-0316323P.
PR   31-AUG-2001; 2001CA-02356540.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wallace DC,  Levy S,  Kerstann K,  Procaccio V;
XX
DR   WPI; 2003-300821/29.
XX
PT   Array containing probes for genes involved in mitochondrial biology,
PT   useful for determining mitochondrial biology gene expression profiles for
PT   use in diagnosing pathologies and identifying biochemical pathways.
XX
PS   Claim 2; SEQ ID NO 274; 201pp; English.
XX
CC   The invention relates to a novel array comprising at least two isolated
CC   nucleotide molecules, each molecule having a sequence capable of uniquely
CC   hybridising to a nucleic acid molecule which is an expression product of
CC   a gene involved in mitochondrial biology. The array comprises two or more
CC   isolated nucleic acid molecules or spots, each molecule having a sequence
CC   chosen from sequence of 994 human probes and 2046 mouse probes. An array
CC   of the invention is useful for determining an expression profile of a
CC   mouse or human sample containing nucleic acid, by contacting the array
CC   with the sample under conditions allowing selective hybridisation, and
CC   measuring hybridisation of nucleic acid in the sample to the array to
CC   produce an expression profile. The array is also useful for determining
CC   an expression profile of a first labelled sample containing nucleic acid
CC   relative to a second, differently labelled sample containing nucleic
CC   acid. The second sample is a reference or a standard. An array is useful
CC   for determining an expression profile diagnostic of an energy-metabolism-
CC   related physiological condition. An array of the invention is useful for
CC   determining mitochondrial biology gene expression profiles of organisms,
CC   such as human, mice and closely related species, tissue and organs of
CC   such organisms, which are useful for determining expression profiles
CC   diagnostic of energy metabolism-related physiological conditions,
CC   diagnosing such physiological conditions, identifying biochemical
CC   pathways, genes, and mutations involved in such physiological conditions,
CC   identifying therapeutic agents useful for preventing and/or treating such
CC   physiological conditions, evaluating and/or monitoring the efficacy of
CC   such therapies, and creating and identifying animal models of human
CC   energy metabolism-related physiological conditions. An array is also
CC   useful for defining expression signatures or profiles for mitochondrial
CC   diseases, as well as distinguishing clinical disorders that result from
CC   oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC   apoptosis and aging. An array of the invention contains probes of genes
CC   not previously recognised to participate in mitochondrial biology. The
CC   sequences shown in ADD32231-ADD33223 represent human mitochondrial DNA
CC   clones used to make the probes of the invention. Some sequences are not
CC   present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC   1906, 2408 and 2643.
XX
SQ   Sequence 455 BP; 127 A; 80 C; 63 G; 185 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 9;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-33;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 96

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 36

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         35 TTTGATGTTAATTAATT 19


RESULT 5
ABK86497
ID   ABK86497 standard; cDNA; 996 BP.
XX
AC   ABK86497;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 cDNA.
XX
KW   Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .996
FT                   /*tag=  a
FT                   /product= "Apo-dystrophin-4 protein appearing as
FT                   AAU98739"
FT                   /partial
FT                   /note= "No start or stop codon shown"
FT                   /transl_except= (pos:7. .9,aa:Xaa)
FT                   /transl_except= (pos:49. .51,aa:Xaa)
FT                   /transl_except= (pos:187. .189,aa:Xaa)
FT                   /transl_except= (pos:202. .204,aa:Xaa)
FT                   /transl_except= (pos:217. .219,aa:Xaa)
FT                   /transl_except= (pos:253. .255,aa:Xaa)
FT                   /transl_except= (pos:265. .270,aa:Xaa-Xaa)
FT                   /transl_except= (pos:394. .396,aa:Xaa)
FT                   /transl_except= (pos:547. .549,aa:Xaa)
FT                   /transl_except= (pos:565. .567,aa:Xaa)
FT                   /transl_except= (pos:616. .618,aa:Xaa)
FT                   /transl_except= (pos:649. .651,aa:Xaa)
FT                   /transl_except= (pos:682. .684,aa:Xaa)
FT                   /transl_except= (pos:709. .711,aa:Xaa)
FT                   /transl_except= (pos:790. .792,aa:Xaa)
FT                   /transl_except= (pos:841. .843,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:880. .882,aa:Xaa)
FT                   /transl_except= (pos:907. .909,aa:Xaa)
FT                   /transl_except= (pos:952. .954,aa:Xaa)
FT                   /transl_except= (pos:982. .984,aa:Xaa)
FT                   /transl_except= (pos:988. .990,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98739.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Claim 5; Page 170-172; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the cDNA sequence for human apo-dystrophin-4
CC   containing a plurality of stop codons some of which may be read through
CC   due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 996;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 919

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 979

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        980 TTTGATGTTAATTAATT 996


RESULT 6
ABK86462
ID   ABK86462 standard; cDNA; 1230 BP.
XX
AC   ABK86462;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 full length cDNA.
XX
KW   Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1230
FT                   /*tag=  a
FT                   /product= "Apo-dystrophin-4"
FT                   /partial
FT                   /note= "No start or stop codon shown"
```

```
FT                   /transl_except= (pos:142. .144,aa:Xaa)
FT                   /transl_except= (pos:241. .244,aa:Xaa)
FT                   /transl_except= (pos:282. .284,aa:Xaa)
FT                   /transl_except= (pos:421. .423,aa:Xaa)
FT                   /transl_except= (pos:436. .438,aa:Xaa)
FT                   /transl_except= (pos:451. .453,aa:Xaa)
FT                   /transl_except= (pos:487. .489,aa:Xaa)
FT                   /transl_except= (pos:499. .504,aa:Xaa-Xaa)
FT                   /transl_except= (pos:628. .630,aa:Xaa)
FT                   /transl_except= (pos:781. .783,aa:Xaa)
FT                   /transl_except= (pos:799. .801,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:883. .885,aa:Xaa)
FT                   /transl_except= (pos:916. .920,aa:Xaa)
FT                   /transl_except= (pos:945. .947,aa:Xaa)
FT                   /transl_except= (pos:1024. .1026,aa:Xaa)
FT                   /transl_except= (pos:1075. .1077,aa:Xaa)
FT                   /transl_except= (pos:1084. .1086,aa:Xaa)
FT                   /transl_except= (pos:1114. .1116,aa:Xaa)
FT                   /transl_except= (pos:1141. .1143,aa:Xaa)
FT                   /transl_except= (pos:1186. .1188,aa:Xaa)
FT                   /transl_except= (pos:1216. .1218,aa:Xaa)
FT                   /transl_except= (pos:1222. .1224,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98729.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 4; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the full length cDNA sequence for human apo-
CC   dystrophin-4 containing a plurality of stop codons some of which may be
CC   read through due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 1230 BP; 404 A; 189 C; 259 G; 378 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 137;  DB 6;  Length 1230;
  Best Local Similarity   100.0%;  Pred. No. 3e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1094 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1153

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1154 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1213

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1214 TTTGATGTTAATTAATT 1230


RESULT 7
ABK86463
ID   ABK86463 standard; cDNA; 1234 BP.
XX
AC   ABK86463;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 full cDNA fragment #1.
XX
KW   Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 9; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is a human apo-dystrophin-4 fragment matching a region
CC   of mouse genomic DNA in the Apo-dystrophin-4 gene region
```

```
XX
SQ   Sequence 1234 BP; 405 A; 190 C; 260 G; 379 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 1234;
  Best Local Similarity  100.0%;  Pred. No. 3e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1157

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1217

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1218 TTTGATGTTAATTAATT 1234


RESULT 8
ABK81996/c
ID   ABK81996 standard; DNA; 2691 BP.
XX
AC   ABK81996;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Human dystrophin 3' untranslated region.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Disclosure; Fig 10; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a human dystrophin polynucleotide
CC   sequence used in the creation of the mini-dystrophin peptides of the
CC   invention
XX
SQ   Sequence 2691 BP; 860 A; 448 C; 461 G; 922 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 2691;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 9
ABK82000/c
ID   ABK82000 standard; DNA; 8689 BP.
XX
AC   ABK82000;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding mini-dystrophin protein deltaH2-R19.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Disclosure; Fig 15; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a mini-dystrophin sequence of the
CC   invention
XX
SQ   Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 8689;
  Best Local Similarity  100.0%;  Pred. No. 5.2e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 7972

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 7912
```

```
Qy        121 TTTGATGTTAATTAATT 137
              ||||||||||||||||
Db       7911 TTTGATGTTAATTAATT 7895


RESULT 10
ABK82002/c
ID   ABK82002 standard; DNA; 11443 BP.
XX
AC   ABK82002;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding mini-dystrophin protein deltaR9-R16.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Disclosure; Fig 17; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a mini-dystrophin sequence of the
CC   invention
XX
SQ   Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10785 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 10726

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10725 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 10666

Qy        121 TTTGATGTTAATTAATT 137
              ||||||||||||||||
Db      10665 TTTGATGTTAATTAATT 10649


RESULT 11
ABK81959/c
ID   ABK81959 standard; DNA; 13957 BP.
XX
AC   ABK81959;
XX
DT   13-AUG-2002  (first entry)
XX
DE   cDNA encoding human dystrophin.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Example 2; Fig 1; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a human dystrophin polynucleotide
CC   sequence used in the creation of the mini-dystrophin peptides of the
CC   invention
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 6e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              ||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 12
ABT10904/c
ID   ABT10904 standard; cDNA; 13957 BP.
XX
AC   ABT10904;
XX
DT   04-DEC-2002  (first entry)
XX
```

DE Human breast cancer associated coding sequence SEQ ID NO: 1038.
XX
KW Human; breast specific gene; breast cancer; differential expression;
KW cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US002176.
XX
PR 25-JAN-2001; 2001US-0263757P.
PR 25-APR-2001; 2001US-0286090P.
PR 23-MAY-2001; 2001US-0292517P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR WPI; 2002-674803/72.
XX
PT Diagnosing breast cancer in a patient comprises detecting the level of
PT gene expression in cell or tissue samples, where a differential gene
PT expression is indicative of breast cancer.
XX
PS Claim 1; SEQ ID NO 1038; 260pp + Sequence Listing; English.
XX
CC The present invention relates to methods of diagnosing breast cancer in a
CC patient, which comprise detecting the level of expression in a tissue
CC sample of two or more genes selected from those shown in ABT09867-
CC ABT11112, where a differential expression of the genes indicates breast
CC cancer. The methods are useful in diagnosing, treating, detecting the
CC progression, and in monitoring treatment of breast cancer in patients.
CC The methods are also useful as a screening tool for agents that modulate
CC the onset or progression of breast cancer. The breast cancer genes may be
CC used as diagnostic markers for the prediction or identification of the
CC malignant state of breast tissue, for confirming the type and progression
CC of cancer, and for drug screening and assays. The present sequence is a
CC coding sequence of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub.published_pct_sequences
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

```
  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 6e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163
```

RESULT 13
ABN95786/c
ID ABN95786 standard; DNA; 13957 BP.
XX
AC ABN95786;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2284 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 2284; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

```
  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 6e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163
```

RESULT 14
ABS69900/c
ID ABS69900 standard; DNA; 13957 BP.
XX
AC ABS69900;
XX
DT 21-NOV-2002 (first entry)
XX
DE Human dystrophin gene.
XX
KW Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;

```
KW   adenosine deaminase deficiency; severe combined immune deficiency; PAH;
KW   beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;
KW   low density lipoprotein gene; familial hypercholesterolaemia;
KW   hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;
KW   phenylalanine hydroxylase gene; gene therapy; phenylketonuria;
KW   dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;
KW   human cystic fibrosis transmembrane conductance regulator gene;
KW   antianaemic; antilipaemic; nootropic; cytostatic; dermatological; human;
KW   alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;
KW   lysosomal arylsulfatase A; omithine transcarbamylase; ARSA; OTC; NP;
KW   purin nucleoside phosphorylase; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2002102731-A1.
XX
PD   01-AUG-2002.
XX
PF   12-FEB-2001; 2001US-00782378.
XX
PR   02-OCT-2000; 2000US-0237747P.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI   Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;
XX
DR   WPI; 2002-690619/74.
XX
PT   Producing vector, by introducing vector having nucleotide sequence,
PT   adenovirus inverted terminal repeats and packaging sequence, and adeno-
PT   associated virus terminal repeat, into cell, and culturing cell.
XX
PS   Disclosure; Page 122-128; 191pp; English.
XX
CC   The present invention relates to a new method of producing a vector. The
CC   method involves introducing recombinant vector having nucleotide sequence
CC   (NS) having 5' and 3' end, left and right inverted terminal repeats of
CC   adenovirus flanking NS, adenovirus packaging sequence linked to inverted
CC   terminal repeat, and adeno-associated virus terminal repeat linked to 3'
CC   end of NS, into cell expressing adenovirus early gene lacking from vector
CC   ; and culturing cell to produce another vector. The method is useful for
CC   generating vectors, especially mAd vectors. The method is useful in
CC   transferring nucleotide sequences of interest into a cell, for gene
CC   transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.
CC   The nucleotide sequences are useful for treating diseases associated with
CC   it, i.e. adenosine deaminase gene associated with adenosine deaminase
CC   deficiency with severe combined immune deficiency, beta-chain of
CC   haemoglobin gene associated with beta-thalassaemia and sickle cell
CC   disease, receptor for low density lipoprotein gene associated with
CC   familial hypercholesterolaemia, hypoxanthine-guanine
CC   phosphoribosyltransferase associated with Lesch-Nyhan syndrome,
CC   phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,
CC   dystrophin gene associated with muscular dystrophy, and human cystic
CC   fibrosis transmembrane conductance regulator gene associated with cystic
CC   fibrosis. The present nucleic acid sequence represents a human disease
CC   gene sequence that was used in the methods of the invention
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity  100.0%;  Pred. No. 6e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 15
ABS70403/c
ID   ABS70403 standard; cDNA; 13977 BP.
XX
AC   ABS70403;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human bone remodelling gene #60.
XX
KW   Bone remodelling; osteoporosis; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   US6426186-B1.
XX
PD   30-JUL-2002.
XX
PF   18-JAN-2000; 2000US-00484970.
XX
PR   18-JAN-2000; 2000US-00484970.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA, Volkmuth W, Walker MG;
XX
DR   WPI; 2002-673014/72.
XX
PT   A combination of polynucleotides which are co-expressed with genes known
PT   to be involved in bone remodeling and osteoporosis are useful in an array
PT   for the diagnosis of bone remodeling and osteoporosis associated
PT   disorders.
XX
PS   Claim 1; Col 185-198; 206pp; English.
XX
CC   The invention relates to a combination comprising a number of
CC   substantially purified and isolated polynucleotides which are co-
CC   expressed with genes known to be involved in bone remodelling and
CC   osteoporosis. The invention is used to diagnose disorders associated with
CC   bone remodelling or osteoporosis. ABS70344-ABS70512 represent human bone
CC   remodelling genes of the invention
XX
SQ   Sequence 13977 BP; 4596 A; 2765 C; 3120 G; 3453 T; 0 U; 43 Other;

  Query Match            92.0%;  Score 126;  DB 6;  Length 13977;
  Best Local Similarity  99.3%;  Pred. No. 1.8e-29;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13306 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13247

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAG-CCGT 119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      13246 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCCGT 13187

Qy        120 GTTTGATGTTAATTAATT 137
              ||||||||||||||||||
Db      13186 GTTTGATGTTAATTAATT 13169


Search completed: April  6, 2004, 11:42:37
Job time : 68.812 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:22:40 ; Search time 574.36 Seconds
(without alignments)
10338.454 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	137	100.0	2148	11	G15848	G15848 human STS C
c	2	137	100.0	2563	9	AK129855	AK129855 Homo sapi
c	3	137	100.0	2691	6	AX538618	AX538618 Sequence
c	4	137	100.0	4658	9	BC028720	BC028720 Homo sapi
c	5	137	100.0	8689	6	AX538622	AX538622 Sequence
c	6	137	100.0	11443	6	AX538624	AX538624 Sequence
c	7	137	100.0	13957	6	AX409637	AX409637 Sequence
c	8	137	100.0	13957	6	AX538581	AX538581 Sequence
c	9	137	100.0	13957	9	HUMDYS	M18533 Homo sapien
	10	137	100.0	98056	9	AC006061	AC006061 Homo sapi
c	11	126	92.0	13977	6	AR220819	AR220819 Sequence
c	12	102.8	75.0	13887	4	AF070485	AF070485 Canis fam
c	13	100.6	73.4	13815	6	AX306153	AX306153 Sequence
c	14	100.6	73.4	13815	6	AX538582	AX538582 Sequence
c	15	100.6	73.4	13815	10	MUSDYSA	M68859 Mouse dystr
	16	100.6	73.4	19307	6	AR093392	AR093392 Sequence
	17	100.6	73.4	19307	6	AR142592	AR142592 Sequence
c	18	100.6	73.4	189131	10	AL645477	AL645477 Mouse DNA
c	19	100.6	73.4	272578	2	AC108338	AC108338 Rattus no
	20	100.6	73.4	279539	2	AC114184	AC114184 Rattus no
c	21	68.4	49.9	13575	5	GGDYS	X13369 Chicken mRN
	22	65	47.4	212120	2	AC023414	AC023414 Homo sapi
	23	39.8	29.1	172734	10	AL928772	AL928772 Mouse DNA
	24	35.8	26.1	94665	3	AC099765	AC099765 Caenorhab
c	25	35.2	25.7	142000	9	AP005404	AP005404 Homo sapi
c	26	35.2	25.7	153228	2	AC139129	AC139129 Mus muscu
c	27	35.2	25.7	155949	9	AC091022	AC091022 Homo sapi
	28	35.2	25.7	156863	2	AC069249	AC069249 Homo sapi
	29	35.2	25.7	158682	2	AC023083	AC023083 Homo sapi
	30	35.2	25.7	165365	2	AC015666	AC015666 Homo sapi
	31	35.2	25.7	165789	10	AC111145	AC111145 Mus muscu
	32	35.2	25.7	180573	9	AC027689	AC027689 Homo sapi
c	33	35	25.5	172422	2	BX649289	BX649289 Danio rer
c	34	35	25.5	231578	2	BX548167	BX548167 Danio rer
c	35	34.6	25.3	137678	8	AP005183	AP005183 Oryza sat
c	36	34.6	25.3	138653	8	AP005178	AP005178 Oryza sat
	37	34.6	25.3	240279	2	AC126486	AC126486 Rattus no
	38	34.6	25.3	270456	2	AC111575	AC111575 Rattus no
c	39	34.4	25.1	495	8	AY202075	AY202075 Arabidops
	40	34.4	25.1	50398	3	AC024771	AC024771 Caenorhab
	41	34.4	25.1	106688	8	ATAC011436	AC011436 Arabidops
c	42	34.4	25.1	206217	2	AC006754	AC006754 Caenorhab
c	43	34.2	25.0	112944	9	AC107082	AC107082 Homo sapi
c	44	34.2	25.0	169755	2	AP001199	AP001199 Homo sapi
c	45	34.2	25.0	171366	9	AC130303	AC130303 Homo sapi

ALIGNMENTS

RESULT 1
G15848/c
LOCUS G15848 2148 bp DNA linear STS 19-JAN-1996
DEFINITION human STS CHLC.UTR_01924_M18533.P56108 clone UTR_01924_M18533, sequence tagged site.
ACCESSION G15848
VERSION G15848.1 GI:1161737
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2148)
AUTHORS Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
TITLE Cooperative Human Linkage Center

```
  JOURNAL   Unpublished (1995)
COMMENT     Synonyms: UTR_01924_M18533, CHLC.UTR_01924_M18533.T36152
            Contact: Dr. Jeffrey C. Murray
            UofI
            The University of Iowa
            Department of Pediatrics, Iowa City, IA 52242, USA
            Tel: (319) 356-3508
            Fax: (319) 356-3347
            Email: jeff-murray@uiowa.edu

            Primer A: AACGCATTTTGGGTTGTTTA
            Primer B: GATATCACGCCAAAAGGATG
            STS size: 189
            PCR Profile:
                        denature:    30 seconds at 94 degrees C
                        annealing:   75 seconds at 55 degrees C
                        extension:   15 seconds at 72 degrees C
                        PCR cycles:  27
                        extension:    6 minutes at 72 degress C
            Protocol:
                        Template:         30ng genomic DNA
                        Primer:           each 1.5 pmole
                        dNTPs:            each 200 uM
                        Taq Polymerase:   0.3 units
                        Total Vol:        10 ul

            Buffer:
                        MgCl2:  1.5mM
                        KCl:     50mM
                        Tris:    10mM
                        pH:     8.3

                        Prepared with primer pairs derived from M18533.
FEATURES              Location/Qualifiers
     source           1. .2148
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
     STS              350. .538
     primer_bind      350. .369
     primer_bind      complement(519. .538)
ORIGIN

  Query Match             100.0%;  Score 137;  DB 11;  Length 2148;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 2
AK129855/c
LOCUS       AK129855               2563 bp    mRNA    linear   PRI 10-SEP-2003
DEFINITION  Homo sapiens cDNA FLJ26345 fis, clone HRT03668.
ACCESSION   AK129855
VERSION     AK129855.1  GI:34526478
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Suzuki,Y.,
            Hata,H., Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Nishikawa,T.,
            Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2563)
  AUTHORS   Sugano,S. and Suzuki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure, Human Genome
            Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction and 5'-end one pass sequencing: Institute of Medical
            Science, University of Tokyo, Laboratory of Genome Structure, Human
            Genome Center; 3'-end one pass sequencing: RAB; clone selection for
            full insert sequencing: RAB and Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1. .2563
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="HRT03668"
                      /tissue_type="heart"
                      /clone_lib="HRT"
                      /note="cloning vector: pME18SFL3"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 9;  Length 2563;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1905 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1846

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1845 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1786

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1785 TTTGATGTTAATTAATT 1769


RESULT 3
AX538618/c
LOCUS       AX538618               2691 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 38 from Patent WO0229056.
ACCESSION   AX538618
VERSION     AX538618.1  GI:25271161
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 38 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES              Location/Qualifiers
     source           1. .2691
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 6;  Length 2691;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-28;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 4
BC028720/c
LOCUS       BC028720                4658 bp    mRNA    linear   PRI 25-AUG-2003
DEFINITION  Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker
            types), transcript variant Dp71b, mRNA (cDNA clone IMAGE:4822807),
            complete cds.
ACCESSION   BC028720
VERSION     BC028720.1  GI:20379675
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4658)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4658)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 46 Row: j Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 5032296
            This clone has the following problem: The cds is short compared to
            the longest cds in the locus.
FEATURES             Location/Qualifiers
     source          1. .4658
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4822807"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     gene            1. .4658
                     /gene="DMD"
                     /note="synonyms: BMD, DXS142, DXS164, DXS206, DXS230,
                     DXS239, DXS268, DXS269, DXS270, DXS272"
                     /db_xref="LocusID:1756"
                     /db_xref="MIM:300377"
     CDS             130. .2037
                     /codon_start=1
                     /product="dystrophin, Dp71b isoform"
                     /protein_id="AAH28720.1"
                     /db_xref="GI:20379676"
                     /db_xref="LocusID:1756"
                     /translation="MREQLKGHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKL
                     RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVN
                     VPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGF
                     CDQRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMR
                     LEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSG
                     RVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLE
                     GDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSIS
                     PNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENR
                     NLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARM
                     QILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGS
                     QTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADDLGRAMESL
                     VSVMTDEEGAE"
     misc_feature    844. .981
                     /note="ZZ; Region: Zinc finger, ZZ type. Zinc finger
                     present in dystrophin, CBP/p300. ZZ in dystrophin binds
                     calmodulin Putative zinc finger"
                     /db_xref="CDD:pfam00569"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 9;  Length 4658;
  Best Local Similarity   100.0%;  Pred. No. 7e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3984 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 3925

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3924 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 3865

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       3864 TTTGATGTTAATTAATT 3848


RESULT 5
AX538622/c
LOCUS       AX538622                8689 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 42 from Patent WO0229056.
ACCESSION   AX538622
VERSION     AX538622.1  GI:25271171
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 42 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .8689
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match           100.0%;  Score 137;  DB 6;  Length 8689;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 7972

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 7912

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       7911 TTTGATGTTAATTAATT 7895


RESULT 6
AX538624/c
LOCUS       AX538624               11443 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 44 from Patent WO0229056.
ACCESSION   AX538624
VERSION     AX538624.1  GI:25271175
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 44 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .11443
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match           100.0%;  Score 137;  DB 6;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10785 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 10726

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10725 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 10666

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      10665 TTTGATGTTAATTAATT 10649


RESULT 7
AX409637/c
LOCUS       AX409637               13957 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2284 from Patent WO0229103.
ACCESSION   AX409637
VERSION     AX409637.1  GI:21442342
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2284 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M18533"
ORIGIN

  Query Match           100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 6e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 8
AX538581/c
LOCUS       AX538581               13957 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 1 from Patent WO0229056.
ACCESSION   AX538581
VERSION     AX538581.1  GI:25271086
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 1 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 6e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180
```

```
Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 9
HUMDYS/c
LOCUS       HUMDYS                 13957 bp    mRNA    linear   PRI 25-MAY-2000
DEFINITION  Homo sapiens dystrophin (DMD) mRNA, complete cds.
ACCESSION   M18533 M17154 M18026 M20250
VERSION     M18533.1  GI:181856
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1699)
  AUTHORS   Koenig,M., Hoffman,E.P., Bertelson,C.J., Monaco,A.P., Feener,C. and
            Kunkel,L.M.
  TITLE     Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
            preliminary genomic organization of the DMD gene in normal and
            affected individuals
  JOURNAL   Cell 50 (3), 509-517 (1987)
  MEDLINE   87273512
   PUBMED   3607877
REFERENCE   2  (bases 1678 to 3830)
  AUTHORS   Hoffman,E.P., Monaco,A.P., Feener,C.C. and Kunkel,L.M.
  TITLE     Conservation of the Duchenne muscular dystrophy gene in mice and
            humans
  JOURNAL   Science 238 (4825), 347-350 (1987)
  MEDLINE   88018015
   PUBMED   3659917
REFERENCE   3  (bases 1 to 13957)
  AUTHORS   Koenig,M., Monaco,A.P. and Kunkel,L.M.
  TITLE     The complete sequence of dystrophin predicts a rod-shaped
            cytoskeletal protein
  JOURNAL   Cell 53 (2), 219-226 (1988)
  MEDLINE   88194521
   PUBMED   3282674
COMMENT     On May 25, 2000 this sequence version replaced gi:340693.
            Draft entry and computer-readable sequence kindly provided by
            M.Koenig, 01-APR-1988 The severity of muscular dystrophy is
            determined by the size of the deleted DNA segment. Deletions found
            in different patients were from positions 302-2200, 473-1168,
            1691-1810, and 1169-3011.
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp21.3-p21.1"
                     /tissue_type="muscle"
                     /dev_stage="fetus"
                     /note="G00-119-850"
     gene            1. .13957
                     /gene="DMD"
     CDS             209. .11266
                     /gene="DMD"
                     /codon_start=1
                     /product="dystrophin"
                     /protein_id="AAA53189.1"
                     /db_xref="GI:181857"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDV
                     DTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQ
                     QITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL
                     MESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTAH
                     QGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLM
                     DLQNQKLKELNDWLTKTEERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVN
                     SLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTE
                     EQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSLK
                     QDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVM
                     ETVTTVTTREQILVKHAQEELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEA
                     VLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSAQALVEQMVNEGVNA
                     DSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQ
                     PTTPSEPTAIKSQLKICKDEVNRLSGLQPQIERLKIQSIALKEKGQGPMFLDADFVAF
                     TNHFKQVFSDVQAREKELQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDY
                     EIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWK
                     KLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL
                     KQCRLLVSDIQTIQPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVY
                     ARKEALKGGLEKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQKAVEEMKRAK
                     EEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNGKCKTLE
                     EVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNP
                     NQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETEKS
                     LHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEA
                     AQRVLSQIDVAQKKLQDVSMKFRLFQKPANFELRLQESKMILDEVKMHLPALETKSVE
                     QEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHY
                     NELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSE
                     VAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIAVTSRAE
                     EWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELND
                     IRPKVDSTRDQAANLMANRGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELE
                     QFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQRGDNLQQRITDER
                     KREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKL
                     ASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWRE
                     IESKFAQFRRLNFAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQ
                     LLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSGRIDIIHSKKTAALQSATPVERVKL
                     QEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI
                     PENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNL
                     RWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKE
                     KLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKLKQTNLQWIKVS
                     RALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFD
                     VQETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRA
                     KQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWT
                     ELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLEELITAAQNLK
                     NKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVL
                     GQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSAD
                     DTRKVHMITENINASWRSIHKRVSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTA
                     NVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQKILRSLEGSD
                     DAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDEL
                     SRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQ
                     EPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNLHSADWQRKIDETLERLQELQEATD
                     ELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQL
                     TTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP
                     WERAISPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKA
                     LCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVD
                     MCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRL
                     GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSM
                     VWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGH
                     KMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
                     PVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESID
                     DEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEY
                     DRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDH
                     NKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSDSM
                     GEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM"

ORIGIN

  Query Match           100.0%;  Score 137;  DB 9;  Length 13957;
  Best Local Similarity  100.0%;  Pred. No. 6e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163
```

RESULT 10
AC006061
LOCUS AC006061 98056 bp DNA linear PRI 01-MAY-2002
DEFINITION Homo sapiens X BAC GSHB-19O24 (Genome Systems Human BAC Library) complete sequence.
ACCESSION AC006061
VERSION AC006061.1 GI:4204246
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 98056)
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L., Kondejewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 7 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 8 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 30, 1999 this sequence version replaced gi:4176317.
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

FEATURES Location/Qualifiers
source 1. .98056
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="GSHB-19O24"
misc_feature 1. .1999
/note="overlaps bases 151178. .153177 of clone AC117405"
/function="clone overlap"
repeat_region 40. .82
/rpt_family="L2"
repeat_region 378. .507
/rpt_family="L1MD3"
repeat_region complement(518. .570)
/rpt_family="(GGA)n"
repeat_region 571. .729
/rpt_family="L1MC4"
repeat_region 730. .987
/rpt_family="MSTD"
repeat_region complement(991. .1400)
/rpt_family="L2"
repeat_region complement(1649. .1779)
/rpt_family="MIR"
repeat_region complement(1783. .2212)
/rpt_family="LTR16A"
repeat_region complement(3626. .3656)
/rpt_family="(GA)n"
repeat_region complement(4713. .5074)
/rpt_family="THE1B"
repeat_region 5702. .7108
/rpt_family="L2"
repeat_region complement(8670. .8734)
/rpt_family="MLT1G"
repeat_region complement(9069. .9144)
/rpt_family="MLT1G"
repeat_region complement(9254. .9351)
/rpt_family="MIR"
repeat_region complement(9537. .9688)
/rpt_family="MER5A"
repeat_region 9797. .9942
/rpt_family="MIR"
repeat_region complement(10403. .10429)
/rpt_family="AT_rich"
repeat_region complement(10673. .10705)
/rpt_family="(CA)n"
repeat_region complement(11460. .11755)
/rpt_family="AluSx"

```
     repeat_region   11977. .12125
                     /rpt_family="MER5A"
     repeat_region   12556. .13643
                     /rpt_family="L1MD2"
     repeat_region   13618. .13808
                     /rpt_family="L1MD3"
     repeat_region   complement(13997. .14297)
                     /rpt_family="MER4D"
     repeat_region   14327. .14625
                     /rpt_family="AluSg"
     repeat_region   complement(14648. .14915)
                     /rpt_family="MER4D"
     repeat_region   complement(14915. .15147)
                     /rpt_family="MER4D"
     repeat_region   15286. .15404
                     /rpt_family="FLAM_C"
     repeat_region   16169. .16463
                     /rpt_family="AluSx"
     repeat_region   16520. .16570
                     /rpt_family="L2"
     repeat_region   16688. .16904
                     /rpt_family="L2"
     repeat_region   complement(16955. .17147)
                     /rpt_family="MIR"
     repeat_region   complement(17171. .17467)
                     /rpt_family="AluY"
     repeat_region   complement(20101. .20241)
                     /rpt_family="MIR"
     STS             22625. .22788
                     /standard_name="DXS1097"
                     /db_xref="GDB:190032"
     repeat_region   24144. .24435
                     /rpt_family="AluSq"
     repeat_region   complement(25010. .25235)
                     /rpt_family="L2"
     repeat_region   25643. .25771
                     /rpt_family="MER58A"
     repeat_region   complement(26111. .26429)
                     /rpt_family="L2"
     repeat_region   complement(26578. .26636)
                     /rpt_family="(CATA)n"
     repeat_region   complement(26956. .27027)
                     /rpt_family="L1MA9"
     repeat_region   complement(27038. .27188)
                     /rpt_family="L2"
     repeat_region   27203. .27641
                     /rpt_family="L1MA8"
     repeat_region   27642. .28179
                     /rpt_family="L1PA4"
     repeat_region   28194. .28495
                     /rpt_family="AluSx"
     repeat_region   complement(28509. .28660)
                     /rpt_family="L2"
     repeat_region   28665. .28830
                     /rpt_family="AluJb"
     repeat_region   complement(28848. .29503)
                     /rpt_family="L2"

  Query Match            100.0%;  Score 137;  DB 9;  Length 98056;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-28;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91659 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 91718

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91719 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 91778

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      91779 TTTGATGTTAATTAATT 91795
```

```
RESULT 11
AR220819/c
LOCUS       AR220819               13977 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 60 from patent US 6426186.
ACCESSION   AR220819
VERSION     AR220819.1  GI:23327696
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 13977)
  AUTHORS   Jones,K.A., Volkmuth,W. and Walker,M.G.
  TITLE     Bone remodeling genes
  JOURNAL   Patent: US 6426186-A 60 30-JUL-2002;
FEATURES             Location/Qualifiers
     source          1. .13977
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            92.0%;  Score 126;  DB 6;  Length 13977;
  Best Local Similarity  99.3%;  Pred. No. 6.9e-25;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13306 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13247

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAG-CCGT 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| 
Db      13246 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCCGT 13187

Qy        120 GTTTGATGTTAATTAATT 137
              ||||||||||||||||||
Db      13186 GTTTGATGTTAATTAATT 13169
```

```
RESULT 12
AF070485/c
LOCUS       AF070485               13887 bp    mRNA    linear   MAM 09-DEC-1998
DEFINITION  Canis familiaris dystrophin mRNA, complete cds.
ACCESSION   AF070485
VERSION     AF070485.1  GI:3982750
KEYWORDS    .
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 13887)
  AUTHORS   Carville,K.S., Mann,C.J., Schatzberg,S.J. and Wilton,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) ANRI, Pathology, University of Western
            Australia, Verdun Street, Nedlands, WA 6018, Australia
FEATURES             Location/Qualifiers
     source          1. .13887
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /strain="breed Golden retriever"
                     /db_xref="taxon:9615"
     5'UTR           <1. .289
     CDS             290. .11332
                     /codon_start=1
                     /product="dystrophin"
                     /protein_id="AAC83646.1"
                     /db_xref="GI:3982751"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQKNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIAKYQLGIEKLLDPEDV
                     ATTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPSQVTREEHFQIHHQMHYSQ
                     QITVSLAQGYERAPSFPKPRFKSYAYTQAAYVTTSDPTRSPLPSQHLETPEDKSFGRS
```

```
                     LTETEANLDSYQTALEEVLSWLLSAEDALQAQGEISNDVEEVKEQFHTHEGYMMDLTS
                     HQGRVGNVLQLGSQLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHKVL
                     MDLQNQQLKELNDWLTKTEERTRKMEKEPLGPDIEDLKRQVQQHKVLQEDLEQEQVRV
                     NSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRFT
                     EEQCLFSAWLSEKEDAVNKIHTTGFKDQSEVLSNLQKLAVLKTDLEKKKQTMDKLCSL
                     NQDLLSALKNTVVAHKMEAWLDNFAQRWDNLVQKLEKSSAQISQAVTTTQPSLTQTTV
                     METVTMVTTREHILVKHAQEELPPPPPQKKRQIIVDSEIRKRLDVDITELHSWITRSE
                     AVLQSPEFAIYRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSAQALVEQMVNEGVN
                     ADSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIITFYNQLQQLEQMTTTAENWLKT
                     QPTTTSEPTAIKSQLKICKDEINRLSALQPQIERLKIQSIALKEKGQGPMFLDADFVA
                     FTNHFNQVFADVQAREKELQTIFDSLPPMRYQETMSTILTWIQQSETKLSIPQVTVTE
                     YDIMEQRLGELQALQSSLQEQQNGLNYLSTTVKEMSKKAPLSDISRKYQSEFEEIEGR
                     WKKLSSQLVEHCQKLEEQMAKLRKIQNHIKTLKKWITEVDVFLKEEWPALGDSEILKR
                     QLKQCRLLVNDIQTIQPSLNSVNEGAQKMKNEAEPEFAGRLETELRELNTQWDYMCRQ
                     VYARKEALKGGLDKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQTAVEEMKR
                     AKEEAQQKEAKVKLLTESVNSVIAQAPPAAQEALKKELDTLTTNYQWLCTRLNGKCKT
                     LEEVWACWHELLSYLEKANKWLSEVEVKLKTTENISGGAEEIAEVLDSLENLMQHSED
                     NPNQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQEIE
                     KSLHLIQESLSSIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGK
                     ETAQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKS
                     VEQEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKL
                     HYNELGAKVTERKQQLEKCLKLSRKMRKEMNALTEWLAATDMELTKRSAVEGMPSNLD
                     SEVAWGKATQKEIEKQKVHLKSVTEVGEALKTVLGKKEMLVEDKLSLLNSNWIAVTSR
                     AEEWLNLLLEYQKHMETFDQNVDYITNWIIQADALLDESEKKKPQQKEDILKRLKAEM
                     NDIRPKVDSTRDQAANLMANRGDHCRKVVEPKISELNHRFAAISHRIKTGKASIPLKE
                     LEQFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMSEDNEGTVKELLQRGDNLQQRITD
                     ERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEK
                     KLASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRW
                     REIESKFAQFRRLNFAQIHTVHEESVVAMTEDMPLEISYVPSTYLTEITHVSQALSEV
                     EELLNAPDLCAQDFEDLFKQEESLKNIKDSLQQISGRIDIIHNKKTAALHSATPAERA
                     KLQEALSRLDFQWERVNNMYKDRQGRFDRSVEKWRRFHYDMKILNQWLTEAEQFLKKT
                     QIPENWEHAKYKWYLKELQDGIGQRQSVVRVLNATGEEIIQQSSKTDASILQEKLGSL
                     NLRWQEVCKQLAERKKRLEEQKNILSEFQRDVNEFVLWLEEADNVANIPLEPGNEQQL
                     KEKLEQVKLLAEELPLRQGILKQLNETGGTVLVSAPLSPEEQDKLENKLKQTNLQWIK
                     VSRNLPEKQEEIEAHVKDLGQLEEQLNHLLLWLSPIRNQLEIYNQPNQTGPFDIKEIE
                     VAVQAKQPDVEGILSKGQHLYKEKPATQPAKRKLEDLSSDWKVVTQLLQELRAKQPGP
                     APGLTTVRAPPSQTVTLVTQPAVTKETAISKLEMPSSLLLEVPALADFNRAWTELTDW
                     LSLLDRVIKSQRVMVGDLEDINEMIIKQKATLQDLEQRRPQLEELITAAQNLKNKTSN
                     QEARTIITDRIERIQSQWDEVQEHLQNRRLQLTEMLKDSTQWLEAKEEAEQVLGQARA
                     KLESWKEAPYTVDAIQKKITETKQLAKDLRQWQINVDVANDLALKLLRDYSADDTRKV
                     HMITENINASWASIHKRLSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTANVLQD
                     ATHKERLLEDSKGVRELMKQWQDLQGEIEAHTDIYHNLDENGQKVLRSLEGSDDAALL
                     QRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAP
                     IGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPREL
                     PPEERAQNVTRLLRKQAEEVNTQWEKLNVHSADWQRKIDEALERLQELQEATDELDLK
                     LRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEITPLKENVSYVNDLARQLTTLGI
                     QLSPYNLNTLEDLNTRWKLLQVAIEDRIRQLHEAHRDFGPASQHFLSTSVQGPWERAI
                     SPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDL
                     LSLSAACDALDQHNLKQNDQPMDILQVINCLTTIYDRLEQEHNNLVNVPLCVDMCLNW
                     LLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLLH
                     DSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPV
                     LHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYP
                     MVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMETPVTLI
                     NFWPVDSAPASSPQLSHDDTHSRIEHYASRLKKMENSNGSYLNDSISPNESIDDEHLL
                     IQHYWRSLNQESPLSQPRSPAQILISLESEERGELERILADLEGRNRNLQAEYDRLKQ
                     QHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLE
                     SQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSESMGEEDL
                     LSPPQDTSTGLEEVMEQLNHSFPSSRGRNTPGKPMREDTM"
     3'UTR           11333. .>13887
     repeat_region   11789. .11828
                     /rpt_type=tandem
                     /rpt_unit="ca"
ORIGIN

  Query Match            75.0%;  Score 102.8;  DB 4;  Length 13887;
  Best Local Similarity  86.9%;  Pred. No. 2e-18;
  Matches  113;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          4 ATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCA 63
              | || ||| ||| |   | ||  |||||||||| |||||||||||| ||||||||||||
Db      13264 AAAATGGAGAAAAATGTTCACATGATGGACAAGTAGTGAAGCTGTGAAATCAGGTGTGCA 13205

Qy         64 CAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTT 123
              |||||| || |||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      13204 CGATTATTAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCATGTTG 13145

Qy        124 GATGTTAATT 133
              ||||||||||
Db      13144 GATGTTAATT 13135


RESULT 13
AX306153/c
LOCUS       AX306153               13815 bp    DNA     linear   PAT 11-DEC-2001
DEFINITION  Sequence 904 from Patent WO0188188.
ACCESSION   AX306153
VERSION     AX306153.1  GI:17645441
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
  TITLE     Method for examining ischemic conditions
  JOURNAL   Patent: WO 0188188-A 904 22-NOV-2001;
            School Juridical Person Nihon University (JP)
FEATURES             Location/Qualifiers
     source          1. .13815
                     /organism="Mus musculus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10090"
ORIGIN

  Query Match            73.4%;  Score 100.6;  DB 6;  Length 13815;
  Best Local Similarity  85.5%;  Pred. No. 8.2e-18;
  Matches  112;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          4 ATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCA 63
              | ||||||||  |       || |||  |||||||||||||| |||||| | |||||||||
Db      13193 AAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAAATTAGGTGTGCA 13134

Qy         64 CAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTT 123
              ||||||  || |||||||||||||||||||||||||||||||||||||||||| ||||||
Db      13133 CAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCTGTGTTT 13074

Qy        124 GATGTTAATTA 134
              ||||||||||
Db      13073 GATGTTAATTA 13063


RESULT 14
AX538582/c
LOCUS       AX538582               13815 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 2 from Patent WO0229056.
ACCESSION   AX538582
VERSION     AX538582.1  GI:25271088
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 2 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .13815
                     /organism="Mus musculus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10090"
ORIGIN

  Query Match            73.4%;  Score 100.6;  DB 6;  Length 13815;
  Best Local Similarity  85.5%;  Pred. No. 8.2e-18;
```

```
  Matches  112;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          4 ATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCA 63
              | |||||||| |      || |||  |||||||||||||| |||||| | |||||||||
Db      13193 AAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAAATTAGGTGTGCA 13134

Qy         64 CAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTT 123
              ||||||  || ||||||||||||||||||||||||||||||||||||||||| ||||||
Db      13133 CAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCTGTGTTT 13074

Qy        124 GATGTTAATTA 134
              ||||||||||
Db      13073 GATGTTAATTA 13063


RESULT 15
MUSDYSA/c
LOCUS       MUSDYSA                13815 bp    mRNA    linear   ROD 24-JUN-1996
DEFINITION  Mouse dystrophin mRNA, complete cds.
ACCESSION   M68859
VERSION     M68859.1  GI:1388025
KEYWORDS    dystrophin.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Koenig,M., Hoffman,E.P., Bertelson,C.J., Monaco,A.P., Feener,C. and
            Kunkel,L.M.
  TITLE     Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
            preliminary genomic organization of the DMD gene in normal and
            affected individuals
  JOURNAL   Cell 50 (3), 509-517 (1987)
  MEDLINE   87273512
   PUBMED   3607877
REFERENCE   2  (sites)
  AUTHORS   Hoffman,E.P., Monaco,A.P., Feener,C.C. and Kunkel,L.M.
  TITLE     Conservation of the Duchenne muscular dystrophy gene in mice and
            humans
  JOURNAL   Science 238 (4825), 347-350 (1987)
  MEDLINE   88018015
   PUBMED   3659917
REFERENCE   3  (bases 1 to 13815)
  AUTHORS   Bies,R.D., Phelps,S.F., Cortez,M.D., Roberts,R., Caskey,C.T. and
            Chamberlain,J.S.
  TITLE     Human and murine dystrophin mRNA transcripts are differentially
            expressed during skeletal muscle, heart, and brain development
  JOURNAL   Nucleic Acids Res. 20 (7), 1725-1731 (1992)
  MEDLINE   92253376
   PUBMED   1579466
REFERENCE   4  (sites)
  AUTHORS   Im,W.B., Phelps,S.F., Copen,E.H., Adams,E.G., Slighton,J.L. and
            Chamberlain,J.S.
  TITLE     Differential expression of dystrophin isoforms in multiple strains
            of mdx mice
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 13815)
  AUTHORS   Chamberlain,J.S., Pearlman,J.A., Muzny,D.M., Civetello,A.,
            Farwell,N.J., Malek,R., Powaser,P., Reeves,A.A., Lee,C.C. and
            Caskey,C.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-1987)
COMMENT     On Jun 24, 1996 this sequence version replaced gi:1387997.
FEATURES             Location/Qualifiers
     source          1. .13815
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/10"
                     /db_xref="taxon:10090"
                     /map="g6pd and pgk-1 of X"
                     /tissue_type="skeletal muscle"
                     /dev_stage="adult"
                     /tissue_lib="lambda zap and lambda gt11 cDNA libraries"
     gene            1. .13815
                     /gene="dystrophin"
     mRNA            <1. .13815
                     /gene="dystrophin"
                     /note="major muscle isoform"
     exon            <1. .251
                     /gene="dystrophin"
                     /note="cloned region"
                     /number=1
                     /evidence=experimental
     conflict        214
                     /gene="dystrophin"
                     /citation=[1]
                     /replace="t"
     CDS             221. .11257
                     /gene="dystrophin"
                     /note="mutation results in a mild muscular dystrophy
                     analogous to BMD; homologue of human Duchenne/Becker
                     muscular dystrophy locus"
                     /codon_start=1
                     /product="dystrophin major muscle isoform"
                     /protein_id="AAB02797.1"
                     /db_xref="GI:1388028"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWINAQFSKFGKQHIDNLFSDL
                     QDGKRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQKNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKTIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT
                     SSWSDGLALNALIHSHRPDLFDWNSVVSQHSATQRLEHAFNIAKCQLGIEKLLDPEDV
                     ATTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRTSSKVTREEHFQLHHQMHYS
                     QQITVSLAQGYEQTSSSPKPRFKSYAFTQAAYVATSDSTQSPYPSQHLEAPRDKSLDS
                     SLMETEVNLDSYQTALEEVLSWLLSAEDTLRAQGEISNDVEEVKEQFHAHEGFMMDLT
                     SHQGLVGNVLQLGSQLVGKGKLSEDEEAEVQEQMNLLNSRWECLRVASMEKQSKLHKV
                     LMDLQNQKLKELDDWLTKTEERTKKMEEEPFGPDLEDLKCQVQQHKVLQEDLEQEQVR
                     VNSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWIVLQDILLKWQHF
                     TEEQCLFSTWLSEKEDAMKNIQTSGFKDQNEMMSSLHKISTLKIDLEKKKPTMEKLSS
                     LNQDLLSALKNKSVTQKMEIWMENFAQRWDNLTQKLEKSSAQISQAVTTTQPSLTQTT
                     VMETVTMVTTREQIMVKHAQEELPPPPPQKKRQITVDSELRKRLDVDITELHSWITRS
                     EAVLQSSEFAVYRKEGNISDLQEKVNAIAREKAEKFRKLQDASRSAQALVEQMANEGV
                     NAESIRQASEQLNSRWTEFCQLLSERVNWLEYQTNIITFYNQLQQLEQMTTTAENLLK
                     TQSTTLSEPTAIKSQLKICKDEVNRLSALQPQIEQLKIQSLQLKEKGQGPMFLDADFV
                     AFTNHFNHIFDGVRAKEKELQTIFDTLPPMRYQETMSSIRTWIQQSESKLSVPYLSVT
                     EYEIMEERLGKLQALQSSLKEQQNGFNYLSDTVKEMAKKAPSEICQKYLSEFEEIEGH
                     WKKLSSQLVESCQKLEEHMNKLRKFQNHIKTLQKWMAEVDVFLKEEWPALGDAEILKK
                     QLKQCRLLVGDIQTIQPSLNSVNEGGQKIKSEAELEFASRLETELRELNTQWDHICRQ
                     VYTRKEALKAGLDKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQTAVEEMKR
                     AKEEALQKETKVKLLTETVNSVIAHAPPSAQEALKKELETLTTNYQWLCTRLNGKCKT
                     LEEVWACWHELLSYLEKANKWLNEVELKLKTMENVPAGPEEITEVLESLENLMHHSEE
                     NPNQIRLLAQTLTDGGVMDELINEELETFNSRWRELHEEAVRKQKLLEQSIQSAQEIE
                     KSLHLIQESLEFIDKQLAAYITDKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGK
                     DANQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLEESKMILDEVKMHLPALETKS
                     VEQEVIQSQLSHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKL
                     HYNELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDTELTKRSAVEGMPSNLD
                     SEVAWGKATQKEIEKQKAHLKSVTELGESLKMVLGKKETLVEDKLSLLNSNWIAVTSR
                     VEEWLNLLLEYQKHMETFDQNIEQITKWIIHADELLDESEKKKPQQKEDILKRLKAEM
                     NDMRPKVDSTRDQAAKLMANRGDHCRKVVEPQISELNRRFAAISHRIKTGKASIPLKE
                     LEQFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMSEDNEGTVNELLQRGDNLQQRITD
                     ERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDEIEK
                     KLASLPEPRDERKLKEIDRELQKKKEELNAVRRQAEGLSENGAAMAVEPTQIQLSKRW
                     RQIESNFAQFRRLNFAQIHTLHEETMVVTTEDMPLDVSYVPSTYLTEISHILQALSEV
                     DHLLNTPELCAKDFEDLFKQEESLKNIKDNLQQISGRIDIIHKKKTAALQSATSMEKV
                     KVQEAVAQMDFQGEKLHRMYKERQGRFDRSVEKWRHFHYDMKVFNQWLNEVEQFFKKT
                     QNPENWEHAKYKWYLKELQDGIGQRQAVVRTLNATGEEIIQQSSKTDVNILQEKLGSL
                     SLRWHDICKELAERRKRIEEQKNVLSEFQRDLNEFVLWLEEADNIAITPLGDEQQLKE
                     QLEQVKLLAEELPLRQGILKQLNETGGAVLVSAPIRPEEQDKLEKKLKQTNLQWIKVS
                     RALPEKQGELEVHLKDFRQLEEQLDHLLLWLSPIRNQLEIYNQPSQAGPFDIKEIEVT
                     VHGKQADVERLLSKGQHLYKEKPSTQPVKRKLEDLRSEWEAVNHLLRELRTKQPDRAP
                     GLSTTGASASQTVTLVTQSVVTKETVISKLEMPSSLLLEVPALADFNRAWTELTDWLS
                     LLDRVIKSQRVMVGDLEDINEMIIKQKATLQDLEQRRPQLEELITAAQNLKNKTSNQE
                     ARTIITDRIERIQIQWDEVQEQLQNRRQQLNEMLKDSTQWLEAKEEAEQVIGQVRGKL
                     DSWKEGPHTVDAIQKKITETKQLAKDLRQRQISVDVANDLALALLLRDYSADDTRKVHM
                     ITENINTSWGNIHKRVSEQEAALEETHRLLQQFPLDLEKFLSWITEAETTANVLQDAS
                     RKEKLLEDSRGVRELMKPWQDLQGEIETHTDIYHNLDENGQKILRSLEGSDEAPLLQR
                     RLDNMNFKWSELQKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPIG
                     GDFPAVQKQNDIHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPRELPP
```

```
                EERAQNVTRLLRKQAEEVNAEWDKLNLRSADWQRKIDEALERLQELQEAADELDLKLR
                QAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVNRVNDLAHQLTTLGIQL
                SPYNLSTLEDLNTRWRLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISP
                NKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDLLS
                LSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVDMCLNWLL
                NVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLLHDS
                IQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLH
                RVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMV
                EYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMETPVTLINF
                WPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESIDDEHLLIQ
                HYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEYDRLKQQH
                EHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLESQ
                LHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSESMGEEDLLS
                PPQDTSTGLEEVMEQLNNSFPSSRGRNAPGKPMREDTM"
   conflict     733
                /gene="dystrophin"
                /citation=[1]
                /replace="a"
   variation    835
                /gene="dystrophin"
                /phenotype="mdx mutant strains 467 and 551"
                /replace="t"
   variation    1324
                /gene="dystrophin"
                /note="the mdx-5cv strain carries this mutation"
                /citation=[4]
                /replace="t"
   conflict     2250
                /gene="dystrophin"
                /citation=[2]
                /replace="t"
   variation    3203
                /gene="dystrophin"
                /phenotype="mdx mutation"
                /replace="t"
   variation    7916
                /gene="dystrophin"
                /note="the mdx-4cv strain carries this mutation"
                /citation=[4]
                /replace="t"
   exon         10423. .10461
                /gene="dystrophin"
                /note="alternatively spliced"
                /number=71
                /evidence=experimental
   exon         10462. .10527
                /gene="dystrophin"
                /note="alternatively spliced"
                /number=72
                /evidence=experimental
   exon         10528. .10593
                /gene="dystrophin"
                /note="alternatively spliced"
                /number=73
                /evidence=experimental
   exon         10594. .10752
                /gene="dystrophin"
                /note="alternatively spliced"
                /number=74
                /evidence=experimental
   exon         11214. .11245
                /gene="dystrophin"
                /note="alternatively spliced"
                /number=78
                /evidence=experimental
   misc_feature 11341. .11343
                /gene="dystrophin"
                /standard_name="embryonic stop codon"
                /phenotype="embryonic isoform"
   polyA_site   13815
                /gene="dystrophin"

Query Match          73.4%;  Score 100.6;  DB 10;  Length 13815;
  Best Local Similarity   85.5%;  Pred. No. 8.2e-18;
  Matches  112;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          4 ATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCA 63
              | |||||||| |       || ||| |||||||||||||| |||||| | |||||||||
Db      13193 AAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAAATTAGGTGTGCA 13134

Qy         64 CAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTT 123
              ||||||  || ||||||||||||||||||||||||||||||||||||||||| ||||||
Db      13133 CAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCTGTGTTT 13074

Qy        124 GATGTTAATTA 134
              ||||||||||
Db      13073 GATGTTAATTA 13063

Search completed: April  6, 2004, 13:02:06
Job time : 581.36 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:33:41 ; Search time 342.077 Seconds
(without alignments)
1500.955 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2470632 seqs, 1873875610 residues

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	137	100.0	137	9	US-09-966-264-1	Sequence 1, Appli
c 2	137	100.0	430	9	US-09-796-692-3505	Sequence 3505, Ap
c 3	137	100.0	430	14	US-10-040-862-3505	Sequence 3505, Ap
c 4	137	100.0	430	15	US-10-057-475B-3505	Sequence 3505, Ap
c 5	137	100.0	430	15	US-10-154-884B-3505	Sequence 3505, Ap
6	137	100.0	996	9	US-09-966-264-2	Sequence 2, Appli
c 7	137	100.0	2691	15	US-10-149-736-38	Sequence 38, Appl
c 8	137	100.0	8689	15	US-10-149-736-42	Sequence 42, Appl
c 9	137	100.0	11443	15	US-10-149-736-44	Sequence 44, Appl
c 10	137	100.0	13957	9	US-09-782-378A-22	Sequence 22, Appl
c 11	137	100.0	13957	9	US-09-880-107-2284	Sequence 2284, Ap
c 12	137	100.0	13957	15	US-10-149-736-1	Sequence 1, Appli
c 13	137	100.0	14069	12	US-10-342-887-434	Sequence 434, App
c 14	137	100.0	14082	12	US-10-342-887-981	Sequence 981, App
c 15	137	100.0	14082	15	US-10-341-434-108	Sequence 108, App
c 16	100.6	73.4	13815	15	US-10-149-736-2	Sequence 2, Appli
c 17	31.6	23.1	292	12	US-10-424-599-138612	Sequence 138612,
c 18	31.6	23.1	663	15	US-10-027-632-199793	Sequence 199793,
c 19	31.6	23.1	2000	9	US-09-938-842A-5066	Sequence 5066, Ap
c 20	31.6	23.1	2000	11	US-09-938-842A-5066	Sequence 5066, Ap
21	31.2	22.8	1977	10	US-09-873-367C-994	Sequence 994, App
22	31.2	22.8	1977	12	US-10-240-425-1268	Sequence 1268, Ap
23	31.2	22.8	1977	12	US-10-342-887-382	Sequence 382, App
24	31.2	22.8	2463	14	US-10-240-965-100	Sequence 100, App
25	30.8	22.5	1157	15	US-10-027-632-208705	Sequence 208705,
26	30.8	22.5	36303	14	US-10-152-724A-24	Sequence 24, Appl
c 27	30.6	22.3	538	9	US-09-867-701-1432	Sequence 1432, Ap
c 28	30.6	22.3	539	9	US-09-917-800A-153	Sequence 153, App
29	30.4	22.2	613	15	US-10-027-632-105562	Sequence 105562,
30	30.4	22.2	613	15	US-10-027-632-126516	Sequence 126516,
31	30.4	22.2	624	15	US-10-027-632-106141	Sequence 106141,
c 32	30.4	22.2	704	15	US-10-027-632-17066	Sequence 17066, A
c 33	30.4	22.2	763	15	US-10-027-632-126517	Sequence 126517,
34	30.4	22.2	786	15	US-10-027-632-126515	Sequence 126515,
35	30	21.9	4974	9	US-09-906-209-17	Sequence 17, Appl
36	30	21.9	6611	10	US-09-934-900-26	Sequence 26, Appl
c 37	30	21.9	7701	15	US-10-459-159-1	Sequence 1, Appli
c 38	30	21.9	7701	15	US-10-427-570A-9	Sequence 9, Appli
c 39	30	21.9	99014	9	US-09-880-107-3428	Sequence 3428, Ap
c 40	29.8	21.8	1015	14	US-10-259-165-589	Sequence 589, App
c 41	29.8	21.8	67191	14	US-10-105-612-1	Sequence 1, Appli
42	29.6	21.6	277	12	US-10-424-599-44445	Sequence 44445, A
43	29.6	21.6	382	12	US-10-424-599-45074	Sequence 45074, A
c 44	29.4	21.5	575	9	US-09-920-300A-4	Sequence 4, Appli
c 45	29.4	21.5	575	13	US-10-033-528-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-966-264-1
; Sequence 1, Application US/09966264
; Patent No. US20020099015A1
; GENERAL INFORMATION:
; APPLICANT: Barber, Elizabeth K
; TITLE OF INVENTION: Gene Expression Control Element DNA
; FILE REFERENCE: 896034605001
; CURRENT APPLICATION NUMBER: US/09/966,264
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/237,079
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 137
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: exon
; LOCATION: (3)..(137)
; OTHER INFORMATION:
; NAME/KEY: polyA_site
; LOCATION: (130)..(135)
; OTHER INFORMATION:
US-09-966-264-1

Query Match 100.0%; Score 137; DB 9; Length 137;
Best Local Similarity 100.0%; Pred. No. 2e-33;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
```

```
Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        121 TTTGATGTTAATTAATT 137
```

RESULT 2
US-09-796-692-3505/c
; Sequence 3505, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-3505

Query Match 100.0%; Score 137; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151
```

RESULT 3
US-10-040-862-3505/c
; Sequence 3505, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-3505

Query Match 100.0%; Score 137; DB 14; Length 430;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151
```

RESULT 4
US-10-057-475B-3505/c
; Sequence 3505, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren

```
;  APPLICANT:  McNeill, Patricia Dianne
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-014402US
;  CURRENT APPLICATION NUMBER: US/10/057,475B
;  CURRENT FILING DATE:  2002-01-22
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 10979
;  SOFTWARE: PastSEQ for Windows Version 3.0
; SEQ ID NO 3505
;    LENGTH: 430
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-057-475B-3505

  Query Match             100.0%;  Score 137;  DB 15;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151


RESULT 5
US-10-154-884B-3505/c
; Sequence 3505, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-013521US
;  CURRENT APPLICATION NUMBER: US/10/154,884B
;  CURRENT FILING DATE:  2002-05-23
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 11290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
;    LENGTH: 430
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-154-884B-3505

  Query Match             100.0%;  Score 137;  DB 15;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151


RESULT 6
US-09-966-264-2
; Sequence 2, Application US/09966264
; Patent No. US20020099015A1
; GENERAL INFORMATION:
;  APPLICANT: Barber, Elizabeth K
;  TITLE OF INVENTION: Gene Expression Control Element DNA
;  FILE REFERENCE: 896034605001
;  CURRENT APPLICATION NUMBER: US/09/966,264
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: US 60/237,079
;  PRIOR FILING DATE: 2000-09-30
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;    LENGTH: 996
;    TYPE: DNA
;    ORGANISM: human
;    FEATURE:
;    NAME/KEY: exon
;    LOCATION: (1)..(996)
;    OTHER INFORMATION:
;    NAME/KEY: misc_feature
;    LOCATION: (710)..(996)
;    OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph
;    OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a
;    OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence
;    OTHER INFORMATION: in the dystrophin DNA
US-09-966-264-2

  Query Match             100.0%;  Score 137;  DB 9;  Length 996;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-33;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 919

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 979

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        980 TTTGATGTTAATTAATT 996


RESULT 7
US-10-149-736-38/c
; Sequence 38, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
;   LENGTH: 2691
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-149-736-38

  Query Match             100.0%;  Score 137;  DB 15;  Length 2691;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 8
US-10-149-736-42/c
; Sequence 42, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 8689
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-42

  Query Match             100.0%;  Score 137;  DB 15;  Length 8689;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-33;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 7972

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 7912

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       7911 TTTGATGTTAATTAATT 7895


RESULT 9
US-10-149-736-44/c
; Sequence 44, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
;   LENGTH: 11443
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-44

  Query Match             100.0%;  Score 137;  DB 15;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10785 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 10726

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10725 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 10666

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      10665 TTTGATGTTAATTAATT 10649


RESULT 10
US-09-782-378A-22/c
; Sequence 22, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
;  APPLICANT: Hearing, Patrick
;  APPLICANT:  Bahou, Wadie
```

```
;  APPLICANT:  Sandalon, Ziv
;  APPLICANT:  Gnatenko, Dmitri
;  TITLE OF INVENTION: Adenoviral Vectors
;  FILE REFERENCE: STONYB-04970
;  CURRENT APPLICATION NUMBER: US/09/782,378A
;  CURRENT FILING DATE:  2001-02-12
;  PRIOR APPLICATION NUMBER: 60/237,747
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 13957
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-782-378A-22

  Query Match             100.0%;  Score 137;  DB 9;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 11
US-09-880-107-2284/c
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
;   LENGTH: 13957
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284

  Query Match             100.0%;  Score 137;  DB 9;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 12
US-10-149-736-1/c
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 13957
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-149-736-1

  Query Match             100.0%;  Score 137;  DB 15;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 13
US-10-342-887-434/c
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
;   LENGTH: 14069
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-342-887-434
```

```
  Query Match             100.0%;  Score 137;  DB 12;  Length 14069;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13411 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13352

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13351 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13292

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13291 TTTGATGTTAATTAATT 13275


RESULT 14
US-10-342-887-981/c
; Sequence 981, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 981
;   LENGTH: 14082
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-342-887-981

  Query Match             100.0%;  Score 137;  DB 12;  Length 14082;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13424 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13365

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13364 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13305

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13304 TTTGATGTTAATTAATT 13288


RESULT 15
US-10-341-434-108/c
; Sequence 108, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 108
;   LENGTH: 14082
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (703)..(11388)
;   OTHER INFORMATION:
US-10-341-434-108

  Query Match             100.0%;  Score 137;  DB 15;  Length 14082;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13424 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13365

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13364 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13305

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13304 TTTGATGTTAATTAATT 13288


Search completed: April  6, 2004, 15:01:39
Job time : 343.077 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 08:36:01 ; Search time 15.2357 Seconds
(without alignments)
4990.154 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	126	92.0	13977	4	US-09-484-970B-60	Sequence 60, Appl
	2	100.6	73.4	19307	3	US-08-836-022A-10	Sequence 10, Appl
	3	100.6	73.4	19307	3	US-09-427-048A-10	Sequence 10, Appl
c	4	28.2	20.6	5326	3	US-09-338-907-124	Sequence 124, App
c	5	28.2	20.6	5326	4	US-09-218-207-124	Sequence 124, App
	6	28	20.4	38844	4	US-09-734-675-3	Sequence 3, Appli
c	7	28	20.4	84495	4	US-09-797-906-3	Sequence 3, Appli
c	8	27.8	20.3	1266	4	US-09-543-681A-2904	Sequence 2904, Ap
c	9	27.8	20.3	2274	3	US-08-492-459-13	Sequence 13, Appl
c	10	27.8	20.3	2274	3	US-08-423-752-13	Sequence 13, Appl
c	11	27.8	20.3	2274	3	US-08-716-873-27	Sequence 27, Appl
c	12	27.8	20.3	2274	3	US-09-368-431-27	Sequence 27, Appl
c	13	27.8	20.3	2274	4	US-09-414-006-13	Sequence 13, Appl
c	14	27.8	20.3	2274	4	US-09-447-223-13	Sequence 13, Appl
	15	27.8	20.3	2799	2	US-08-874-186-42	Sequence 42, Appl
c	16	27.8	20.3	7218	1	US-08-232-463-14	Sequence 14, Appl
c	17	27.6	20.1	1173	4	US-09-107-532A-2915	Sequence 2915, Ap
c	18	27.6	20.1	3705	5	PCT-US96-03940-7	Sequence 7, Appli
	19	27.6	20.1	3786	4	US-08-961-527-182	Sequence 182, App
c	20	27.6	20.1	5648	5	PCT-US96-03940-1	Sequence 1, Appli
	21	27.6	20.1	18355	4	US-08-956-171E-67	Sequence 67, Appl
	22	27.4	20.0	2421	3	US-08-851-843A-51	Sequence 51, Appl
	23	27.4	20.0	2421	3	US-08-974-549A-218	Sequence 218, App
	24	27.4	20.0	2421	3	US-08-854-050-51	Sequence 51, Appl
	25	27.4	20.0	2421	4	US-09-430-323-51	Sequence 51, Appl
	26	27.4	20.0	2421	4	US-09-402-181B-218	Sequence 218, App
	27	27.4	20.0	2421	4	US-09-721-456-218	Sequence 218, App
c	28	27.4	20.0	3459	4	US-09-016-434-1363	Sequence 1363, Ap
	29	27.4	20.0	58407	4	US-08-916-421B-2	Sequence 2, Appli
	30	27.2	19.9	770	4	US-09-205-258-161	Sequence 161, App
	31	27.2	19.9	1296	4	US-09-134-001C-1420	Sequence 1420, Ap
	32	27.2	19.9	2508	2	US-08-850-993-1	Sequence 1, Appli
	33	27.2	19.9	13417	2	US-08-637-759B-37	Sequence 37, Appl
	34	27.2	19.9	13417	3	US-08-871-355A-37	Sequence 37, Appl
	35	27.2	19.9	13417	4	US-09-201-945-37	Sequence 37, Appl
	36	26.8	19.6	888	4	US-09-134-001C-1629	Sequence 1629, Ap
	37	26.8	19.6	1221	4	US-09-134-001C-733	Sequence 733, App
	38	26.8	19.6	5873	1	US-07-928-464-4	Sequence 4, Appli
	39	26.8	19.6	5873	5	PCT-US93-07347-4	Sequence 4, Appli
	40	26.8	19.6	5890	1	US-07-928-464-3	Sequence 3, Appli
	41	26.8	19.6	5890	1	US-07-928-464-5	Sequence 5, Appli
	42	26.8	19.6	5890	1	US-07-928-464-6	Sequence 6, Appli
	43	26.8	19.6	5890	5	PCT-US93-07347-3	Sequence 3, Appli
	44	26.8	19.6	5890	5	PCT-US93-07347-5	Sequence 5, Appli
	45	26.8	19.6	5890	5	PCT-US93-07347-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-484-970B-60/c
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
; NAME/KEY: unsure
; LOCATION: 11721-11761, 12294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match 92.0%; Score 126; DB 4; Length 13977;
Best Local Similarity 99.3%; Pred. No. 4.8e-31;
Matches 137; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13306 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13247

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAG-CCGT 119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      13246 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCCGT 13187

Qy        120 GTTTGATGTTAATTAATT 137
              ||||||||||||||||||
Db      13186 GTTTGATGTTAATTAATT 13169
```

RESULT 2
US-08-836-022A-10
; Sequence 10, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.

```
;    APPLICANT:  Fisher, Krishna J.
;    APPLICANT:  Chen, Shu-Jen
;    APPLICANT:  Weitzman, Matthew
;    TITLE OF INVENTION:  Improved Adenovirus Virus and
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P O Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,022A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/331,381
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GNVPN.008PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19307 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-836-022A-10

  Query Match             73.4%;  Score 100.6;  DB 3;  Length 19307;
  Best Local Similarity   85.5%;  Pred. No. 8e-23;
  Matches  112;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          4 ATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCA 63
              | ||||||||  |      || |||  |||||||||||||| |||||| | |||||||||
Db       1492 AAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAAATTAGGTGTGCA 1551

Qy         64 CAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTT 123
              ||||||  || ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1552 CAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCTGTGTTT 1611

Qy        124 GATGTTAATTA 134
              ||||||||||
Db       1612 GATGTTAATTA 1622


RESULT 3
US-09-427-048A-10
; Sequence 10, Application US/09427048A
; Patent No. 6203975
;   GENERAL INFORMATION:
;        APPLICANT: Trustees of the University of Pennsylvania
;                   Wilson, James M.
;                   Fisher, Krishna J.
;                   Chen, Shu-Jen
;                   Weitzman, Matthew
;        TITLE OF INVENTION: Improved Adenovirus Virus and
;                            Methods of Use Thereof
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Howson and Howson
;             STREET: Spring House Corporate Cntr, P O Box 457
;             CITY: Spring House
;             STATE: Pennsylvania
;             COUNTRY: USA
;             ZIP: 19477
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/427,048A
;             FILING DATE: 21-Oct-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/836,022
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Bak, Mary E.
;             REGISTRATION NUMBER: 31,215
;             REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 215-540-9200
;             TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 10:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 19307 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: unknown
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

  Query Match             73.4%;  Score 100.6;  DB 3;  Length 19307;
  Best Local Similarity   85.5%;  Pred. No. 8e-23;
  Matches  112;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          4 ATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCA 63
              | ||||||||  |      || |||  |||||||||||||| |||||| | |||||||||
Db       1492 AAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAAATTAGGTGTGCA 1551

Qy         64 CAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTT 123
              ||||||  || ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1552 CAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCTGTGTTT 1611

Qy        124 GATGTTAATTA 134
              ||||||||||
Db       1612 GATGTTAATTA 1622


RESULT 4
US-09-338-907-124/c
; Sequence 124, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Ilya, Chumakov
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: PROSTATE CANCER GENE
;  FILE REFERENCE: GENSET.18CP1CP
;  CURRENT APPLICATION NUMBER: US/09/338,907
;  CURRENT FILING DATE: 1999-06-23
;  EARLIER APPLICATION NUMBER: 08/996,306
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: 60/099,658
;  EARLIER FILING DATE: 1998-09-09
;  EARLIER APPLICATION NUMBER: 09/218,207
;  EARLIER FILING DATE: 1998-12-22
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
; SEQ ID NO 124
;   LENGTH: 5324
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 31..588
;   NAME/KEY: polyA_signal
;   LOCATION: 5297..5302
;   OTHER INFORMATION: AATAAA
US-09-338-907-124

  Query Match             20.6%;  Score 28.2;  DB 3;  Length 5326;
  Best Local Similarity   52.8%;  Pred. No. 11;
  Matches   57;  Conservative    1;  Mismatches   50;  Indels    0;  Gaps    0;

Qy         11 AAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTAT 70
              |||||| || |  |   |||:   | ||||  |    |  || |  | | |  | |
Db        661 ACAAAGAGAACACGGAGCTGGR.GCGCGGTGGCACGTGTATGTCTGTAGTCCTAACTACT 602

Qy         71 CAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCG 118
              |||||  ||  |   | | | ||  || || ||  || |||  ||| |
Db        601 CAGGAGGACTGCTTTAGCCACGTTGGGCAGCAAATGCCTGACTAGCTG 554


RESULT 5
US-09-218-207-124/c
; Sequence 124, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Ilya, Chumakov
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Prostate cancer gene
;  FILE REFERENCE: GENSET.018CP1
;  CURRENT APPLICATION NUMBER: US/09/218,207
;  CURRENT FILING DATE: 1998-12-22
;  EARLIER APPLICATION NUMBER: 08/996,306
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: 60/099,658
;  EARLIER FILING DATE: 1998-09-09
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
; SEQ ID NO 124
;   LENGTH: 5324
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 31..588
;   NAME/KEY: polyA_signal
;   LOCATION: 5297..5302
;   OTHER INFORMATION: AATAAA
US-09-218-207-124

  Query Match             20.6%;  Score 28.2;  DB 4;  Length 5326;
  Best Local Similarity   52.8%;  Pred. No. 11;
  Matches   57;  Conservative    1;  Mismatches   50;  Indels    0;  Gaps    0;

Qy         11 AAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTAT 70
              |||||| || |  |   |||:   | |-||| |    |  || |  | | |  | |
Db        661 ACAAAGAGAACACGGAGCTGGR.GCGCGGTGGCACGTGTATGTCTGTAGTCCTAACTACT 602

Qy         71 CAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCG 118
              |||||  ||  |   | | | ||  || || ||  || |||  ||| |
Db        601 CAGGAGGACTGCTTTAGCCACGTTGGGCAGCAAATGCCTGACTAGCTG 554


RESULT 6
US-09-734-675-3
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL000862
;  CURRENT APPLICATION NUMBER: US/09/734,675
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 38844
;   TYPE: DNA
;   ORGANISM: Human
US-09-734-675-3

  Query Match             20.4%;  Score 28;  DB 4;  Length 38844;
  Best Local Similarity   50.8%;  Pred. No. 23;
  Matches   67;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              | |||  | ||    |  ||||    |||  | || || | ||  | | || | |  ||
Db        526 AATATGTATGAGTGAATGAATAGATTAATACATAAATGATAAATTTATTAAATAATTTGA 585

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              |  |||  |   |  | | || | |  |||  |    ||   |  || | ||  | | | |
Db        586 ACGAAATCAATCGAGAGGCACCGAGAATAAATTTGTGTCCTAGAAGTAAGAAGACCTGAG 645

Qy        121 TTTGATGTTAAT 132
              |||||  | | |
Db        646 TTTGAGATAACT 657


RESULT 7
US-09-797-906-3/c
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
;  APPLICANT:  Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
;  TITLE OF INVENTION:  ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL001151CIP
;  CURRENT APPLICATION NUMBER: US/09/797,906
;  CURRENT FILING DATE:  2001-03-05
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 84495
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(84495)
;   OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3

  Query Match             20.4%;  Score 28;  DB 4;  Length 84495;
  Best Local Similarity   55.0%;  Pred. No. 29;
  Matches   55;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          2 TTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTG 61
              ||| |||  ||||| |||| ||  |   |||| | |  | ||| |  ||      |||
Db       9226 TTAAAAAAAAAAAAAAAAAAAAAATTACATGACATGGGAAGCAGCAGGAAAACGTAATGTA 9167

Qy         62 CACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 101
              ||  |  |  |  |||| |    ||| |||  ||   |||
Db       9166 CAATAAAAAAAAAAACAGCAGTCAACAAAACAGAAACAGA 9127


RESULT 8
US-09-543-681A-2904/c
; Sequence 2904, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
```

```
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2904
;   LENGTH: 1266
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-2904

  Query Match             20.3%;  Score 27.8;  DB 4;  Length 1266;
  Best Local Similarity   65.1%;  Pred. No. 9;
  Matches   41;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          65 AATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTTG 124
               | ||||||| |  |  | ||| | |||||  || |||||||     ||| | ||  ||
Db         886 ATTTATCAGAGAACCAGCTAAATAAGAGTGAGCCAGGAATAGCAAACCAAGGCATGAGTG 827

Qy         125 ATG 127
               |||
Db         826 ATG 824


RESULT 9
US-08-492-459-13/c
; Sequence 13, Application US/08492459
; Patent No. 6015689
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  REGULATION OF AUREOBASIDIN SENSITIVITY IN FUNGUS
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/492,459
;      FILING DATE:  June 20, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2274
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-492-459-13

  Query Match             20.3%;  Score 27.8;  DB 3;  Length 2274;
  Best Local Similarity   57.5%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          50 AACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCAT 109
               ||| ||    || | ||  | || | | ||||| ||||   |  || ||| | |
Db        1848 AACACATTAATGTAAAAAAAGAAAGACAAACACAAAAACAAAAAAACTTAAAAAAACTGT 1789

Qy         110 GAGAAGCCGTGTTTGATGTTAATTAAT 136
               |   |  |   |||||| ||||||||
Db        1788 GCCCATTCTATATTGATGGTAATTAAT 1762


RESULT 10
US-08-423-752-13/c
; Sequence 13, Application US/08423752
; Patent No. 6022949
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  A GENE CODING FOR A PROTEIN REGULATING
;    TITLE OF INVENTION:  AUREOBASIDIN SENSITIVITY
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/423,752
;      FILING DATE:  April 18, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2274
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-423-752-13

  Query Match             20.3%;  Score 27.8;  DB 3;  Length 2274;
  Best Local Similarity   57.5%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          50 AACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCAT 109
               ||| ||    || | ||  | || | | ||||| ||||   |  || ||| | |
Db        1848 AACACATTAATGTAAAAAAAGAAAGACAAACACAAAAACAAAAAAACTTAAAAAAACTGT 1789

Qy         110 GAGAAGCCGTGTTTGATGTTAATTAAT 136
               |   |  |   |||||| ||||||||
Db        1788 GCCCATTCTATATTGATGGTAATTAAT 1762


RESULT 11
US-08-716-873-27/c
; Sequence 27, Application US/08716873
```

```
; Patent No. 6194166
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:- GENE REGULATING AUREOBASIDIN SENSITIVITY
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,873
;      FILING DATE:  September 20, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2274 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-716-873-27

  Query Match             20.3%;  Score 27.8;  DB 3;  Length 2274;
  Best Local Similarity   57.5%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          50 AACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCAT 109
               ||| ||    || | ||  | | ||  | | ||||| ||||   |  || ||| |   |
Db        1848 AACACATTAATGTAAAAAAAGAAAGACAAACACAAAAACAAAAAAACTTAAAAAAACTGT 1789

Qy          110 GAGAAGCCGTGTTTGATGTTAATTAAT 136
               |   | |    |||||| ||||||||
Db        1788 GCCCATTCTATATTGATGGTAATTAAT 1762


RESULT 12
US-09-368-431-27/c
; Sequence 27, Application US/09368431
; Patent No. 6294651
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  GENE REGULATING AUREOBASIDIN SENSITIVITY
;    TITLE OF INVENTION:  (AS AMENDED)
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WENDEROTH, LIND & PONACK, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/368,431
;      FILING DATE:  August 5, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/716,873
;      FILING DATE:  September 20, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2274 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-09-368-431-27

  Query Match             20.3%;  Score 27.8;  DB 3;  Length 2274;
  Best Local Similarity   57.5%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          50 AACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCAT 109
               ||| ||    || | ||  | | ||  | | ||||| ||||   |  || ||| |   |
Db        1848 AACACATTAATGTAAAAAAAGAAAGACAAACACAAAAACAAAAAAACTTAAAAAAACTGT 1789

Qy          110 GAGAAGCCGTGTTTGATGTTAATTAAT 136
               |   | |    |||||| ||||||||
Db        1788 GCCCATTCTATATTGATGGTAATTAAT 1762


RESULT 13
US-09-414-006-13/c
; Sequence 13, Application US/09414006
; Patent No. 6348577
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  REGULATION OF AUREOBASIDIN SENSITIVITY (AS
;    TITLE OF INVENTION:  AMENDED)
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/414,006
;      FILING DATE:  October 7, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/492,459
;      FILING DATE:  June 20, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
```

```
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2274
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-09-414-006-13

  Query Match             20.3%;  Score 27.8;  DB 4;  Length 2274;
  Best Local Similarity   57.5%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          50 AACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCAT 109
               ||| ||    || | ||  |  | ||  | | ||||| ||||   |  || ||| |   |
Db        1848 AACACATTAATGTAAAAAAAGAAAGACAAACACAAAAACAAAAAAACTTAAAAAAACTGT 1789

Qy          110 GAGAAGCCGTGTTTGATGTTAATTAAT 136
               |   |  |    |||||| ||||||||
Db        1788 GCCCATTCTATATTGATGGTAATTAAT 1762


RESULT 14
US-09-447-223-13/c
; Sequence 13, Application US/09447223
; Patent No. 6432664
;   GENERAL INFORMATION:
;        APPLICANT: Takashi OKADO et al.
;        TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
;                            AUREOBASIDIN SENSITIVITY
;        NUMBER OF SEQUENCES: 22
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Wenderoth, Lind & Ponack
;             STREET: 805 Fifteenth Street, N.W., #700
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: MS-DOS
;             SOFTWARE: Wordperfect 5.1
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/447,223
;             FILING DATE: 23-No. 6432664-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/243,403
;             FILING DATE: May 16, 1994
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Warren M. Cheek, Jr.
;             REGISTRATION NUMBER: 33,367
;             REFERENCE/DOCKET NUMBER: <Unknown>
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-371-8850
;             TELEFAX: <Unknown>
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 13:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2274
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: genomic DNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-447-223-13

  Query Match             20.3%;  Score 27.8;  DB 4;  Length 2274;
  Best Local Similarity   57.5%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          50 AACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCAT 109
               ||| ||    || | ||  |  | ||  | | ||||| ||||   |  || ||| |   |
Db        1848 AACACATTAATGTAAAAAAAGAAAGACAAACACAAAAACAAAAAAACTTAAAAAAACTGT 1789

Qy          110 GAGAAGCCGTGTTTGATGTTAATTAAT 136
               |   |  |    |||||| ||||||||
Db        1788 GCCCATTCTATATTGATGGTAATTAAT 1762


RESULT 15
US-08-874-186-42
; Sequence 42, Application US/08874186
; Patent No. 5989885
;  GENERAL INFORMATION:
;    APPLICANT:  Teng, David H-F.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Perry III, William L.
;    APPLICANT:  Skolnick, Mark H.
;    TITLE OF INVENTION:  SPECIFIC MUTATIONS OF MAP KINASE KINASE
;    TITLE OF INVENTION:  4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
;    TITLE OF INVENTION:  SUPPRESSOR IN VARIOUS TYPES OF CANCER
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/874,186
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,482
;      FILING DATE:  10-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saxe, Stephen A.
;      REGISTRATION NUMBER:  38,609
;      REFERENCE/DOCKET NUMBER:  24884-121392-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4848
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2799 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1..764
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  765..892
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  893..2799
US-08-874-186-42

  Query Match             20.3%;  Score 27.8;  DB 2;  Length 2799;
```

Best Local Similarity 57.5%; Pred. No. 12;
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy          18 AAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGGAAC 77
               ||||||  ||  | | ||| | |  || ||| |  | | | || | |||| | | |
Db        2082 AAATAAAAAAAAATTAGACATGGAGTATTTGGGAAATGTGATTTACATATTTATTAAGGAT 2141

Qy          78 ACCCCAAAACCAAAGTGAGGTAGAAAT 104
               ||||| | |   || |||  | |  ||
Db        2142 ACCCCGAGAATGAACTGAAATTGTTAT 2168
```

Search completed: April 6, 2004, 14:14:17
Job time : 16.2357 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 15:01:53 ; Search time 61.7891 Seconds
(without alignments)
8309.593 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2470632 seqs, 1873875610 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	137	100.0	137	9	US-09-966-264-1	Sequence 1, Appli
c	2	137	100.0	430	9	US-09-796-692-3505	Sequence 3505, Ap
c	3	137	100.0	430	14	US-10-040-862-3505	Sequence 3505, Ap
c	4	137	100.0	430	15	US-10-057-475B-3505	Sequence 3505, Ap
c	5	137	100.0	430	15	US-10-154-884B-3505	Sequence 3505, Ap
	6	137	100.0	996	9	US-09-966-264-2	Sequence 2, Appli
c	7	137	100.0	2691	15	US-10-149-736-38	Sequence 38, Appl
c	8	137	100.0	8689	15	US-10-149-736-42	Sequence 42, Appl
c	9	137	100.0	11443	15	US-10-149-736-44	Sequence 44, Appl
c	10	137	100.0	13957	9	US-09-782-378A-22	Sequence 22, Appl
c	11	137	100.0	13957	9	US-09-880-107-2284	Sequence 2284, Ap
c	12	137	100.0	13957	15	US-10-149-736-1	Sequence 1, Appli
c	13	137	100.0	14069	12	US-10-342-887-434	Sequence 434, App
c	14	137	100.0	14082	12	US-10-342-887-981	Sequence 981, App
c	15	137	100.0	14082	15	US-10-341-434-108	Sequence 108, App
c	16	42	30.7	13815	15	US-10-149-736-2	Sequence 2, Appli
c	17	22	16.1	22	9	US-09-966-264-13	Sequence 13, Appl
c	18	20	14.6	690	15	US-10-131-827-8211	Sequence 8211, Ap
c	19	19	13.9	598	15	US-10-027-632-293205	Sequence 293205,
c	20	19	13.9	613	15	US-10-027-632-81394	Sequence 81394, A
	21	19	13.9	700	12	US-10-425-114-6882	Sequence 6882, Ap
c	22	19	13.9	42521	10	US-09-373-658-23	Sequence 23, Appl
c	23	19	13.9	42521	11	US-09-989-687-23	Sequence 23, Appl
c	24	19	13.9	188971	12	US-10-235-192A-27	Sequence 27, Appl
	25	18	13.1	18	9	US-09-966-264-22	Sequence 22, Appl
	26	18	13.1	532	15	US-10-027-632-274023	Sequence 274023,
	27	18	13.1	689	15	US-10-027-632-280438	Sequence 280438,
	28	18	13.1	717	12	US-10-282-122A-21513	Sequence 21513, A
	29	18	13.1	765	15	US-10-027-632-148232	Sequence 148232,
	30	18	13.1	806	15	US-10-027-632-154166	Sequence 154166,
	31	18	13.1	806	15	US-10-027-632-154167	Sequence 154167,
	32	18	13.1	806	15	US-10-027-632-154168	Sequence 154168,
	33	18	13.1	806	15	US-10-027-632-154169	Sequence 154169,
c	34	18	13.1	878	9	US-09-770-445-538	Sequence 538, App
c	35	18	13.1	1909	15	US-10-027-632-258397	Sequence 258397,
c	36	18	13.1	1909	15	US-10-027-632-258398	Sequence 258398,
	37	18	13.1	2278	15	US-10-027-632-102777	Sequence 102777,
	38	18	13.1	3415	10	US-09-814-353-19075	Sequence 19075, A
c	39	18	13.1	4001	13	US-10-135-322-34	Sequence 34, Appl
c	40	18	13.1	7028	13	US-10-135-322-33	Sequence 33, Appl
	41	18	13.1	70768	13	US-10-135-322-13	Sequence 13, Appl
c	42	17	12.4	327	12	US-10-085-783A-12712	Sequence 12712, A
c	43	17	12.4	327	15	US-10-242-535A-12712	Sequence 12712, A
	44	17	12.4	355	12	US-10-424-599-396	Sequence 396, App
c	45	17	12.4	457	10	US-09-918-995-13211	Sequence 13211, A

ALIGNMENTS

RESULT 1
US-09-966-264-1
; Sequence 1, Application US/09966264
; Patent No. US20020099015A1
; GENERAL INFORMATION:
; APPLICANT: Barber, Elizabeth K
; TITLE OF INVENTION: Gene Expression Control Element DNA
; FILE REFERENCE: 896034605001
; CURRENT APPLICATION NUMBER: US/09/966,264
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/237,079
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 137
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: exon
; LOCATION: (3)..(137)
; OTHER INFORMATION:
; NAME/KEY: polyA_site
; LOCATION: (130)..(135)
; OTHER INFORMATION:
US-09-966-264-1

Query Match 100.0%; Score 137; DB 9; Length 137;
Best Local Similarity 100.0%; Pred. No. 9.5e-61;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
```

```
Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        121 TTTGATGTTAATTAATT 137
```

RESULT 2
US-09-796-692-3505/c
; Sequence 3505, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-3505

Query Match 100.0%; Score 137; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 1e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151
```

RESULT 3
US-10-040-862-3505/c
; Sequence 3505, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-3505

Query Match 100.0%; Score 137; DB 14; Length 430;
Best Local Similarity 100.0%; Pred. No. 1e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151
```

RESULT 4
US-10-057-475B-3505/c
; Sequence 3505, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren

```
;  APPLICANT:  McNeill, Patricia Dianne
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-014402US
;  CURRENT APPLICATION NUMBER: US/10/057,475B
;  CURRENT FILING DATE:  2002-01-22
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 10979
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-057-475B-3505

  Query Match             100.0%;  Score 137;  DB 15;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 1e-60;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151


RESULT 5
US-10-154-884B-3505/c
; Sequence 3505, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-013521US
;  CURRENT APPLICATION NUMBER: US/10/154,884B
;  CURRENT FILING DATE:  2002-05-23
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 11290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-154-884B-3505

  Query Match             100.0%;  Score 137;  DB 15;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 1e-60;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 228

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 168

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        167 TTTGATGTTAATTAATT 151


RESULT 6
US-09-966-264-2
; Sequence 2, Application US/09966264
; Patent No. US20020099015A1
; GENERAL INFORMATION:
;  APPLICANT: Barber, Elizabeth K
;  TITLE OF INVENTION: Gene Expression Control Element DNA
;  FILE REFERENCE: 896034605001
;  CURRENT APPLICATION NUMBER: US/09/966,264
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: US 60/237,079
;  PRIOR FILING DATE: 2000-09-30
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 996
;   TYPE: DNA
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1)..(996)
;   OTHER INFORMATION:
;   NAME/KEY: misc_feature
;   LOCATION: (710)..(996)
;   OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph
;   OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a
;   OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence
;   OTHER INFORMATION: in the dystrophin DNA
US-09-966-264-2

  Query Match             100.0%;  Score 137;  DB 9;  Length 996;
  Best Local Similarity   100.0%;  Pred. No. 1e-60;
```

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 919

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 979

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        980 TTTGATGTTAATTAATT 996
```

RESULT 7
US-10-149-736-38/c
; Sequence 38, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 2691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-38

Query Match 100.0%; Score 137; DB 15; Length 2691;
Best Local Similarity 100.0%; Pred. No. 1.1e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897
```

RESULT 8
US-10-149-736-42/c
; Sequence 42, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 8689
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-42

Query Match 100.0%; Score 137; DB 15; Length 8689;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 7972

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 7912

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       7911 TTTGATGTTAATTAATT 7895
```

RESULT 9
US-10-149-736-44/c
; Sequence 44, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 11443
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-44

Query Match 100.0%; Score 137; DB 15; Length 11443;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10785 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 10726

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10725 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 10666

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      10665 TTTGATGTTAATTAATT 10649
```

RESULT 10
US-09-782-378A-22/c
; Sequence 22, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie

; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-782-378A-22

Query Match 100.0%; Score 137; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy 121 TTTGATGTTAATTAATT 137
|||||||||||||||||
Db 13179 TTTGATGTTAATTAATT 13163

RESULT 11
US-09-880-107-2284/c
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284

Query Match 100.0%; Score 137; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy 121 TTTGATGTTAATTAATT 137
|||||||||||||||||
Db 13179 TTTGATGTTAATTAATT 13163

RESULT 12
US-10-149-736-1/c
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-1

Query Match 100.0%; Score 137; DB 15; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy 121 TTTGATGTTAATTAATT 137
|||||||||||||||||
Db 13179 TTTGATGTTAATTAATT 13163

RESULT 13
US-10-342-887-434/c
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
; LENGTH: 14069
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-434

Query Match 100.0%; Score 137; DB 12; Length 14069;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13411 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13352

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13351 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13292

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13291 TTTGATGTTAATTAATT 13275
```

RESULT 14
US-10-342-887-981/c
; Sequence 981, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 981
; LENGTH: 14082
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-981

Query Match 100.0%; Score 137; DB 12; Length 14082;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13424 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13365

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13364 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13305

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13304 TTTGATGTTAATTAATT 13288
```

RESULT 15
US-10-341-434-108/c
; Sequence 108, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 108
; LENGTH: 14082
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (703)..(11388)
; OTHER INFORMATION:
US-10-341-434-108

Query Match 100.0%; Score 137; DB 15; Length 14082;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13424 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13365

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13364 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13305

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13304 TTTGATGTTAATTAATT 13288
```

Search completed: April 6, 2004, 17:51:12
Job time : 63.7891 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:14:32 ; Search time 15.1147 Seconds
(without alignments)
5030.075 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	117	85.4	13977	4	US-09-484-970B-60	Sequence 60, Appl
	2	42	30.7	19307	3	US-08-836-022A-10	Sequence 10, Appl
	3	42	30.7	19307	3	US-09-427-048A-10	Sequence 10, Appl
	4	19	13.9	4766	5	PCT-US93-07261-10	Sequence 10, Appl
	5	18	13.1	750	4	US-09-107-532A-866	Sequence 866, App
	6	18	13.1	889	2	US-08-935-886-7	Sequence 7, Appli
	7	18	13.1	889	2	US-08-935-886-13	Sequence 13, Appl
	8	18	13.1	2076	4	US-09-134-001C-1838	Sequence 1838, Ap
c	9	17	12.4	92	1	US-07-839-751-1	Sequence 1, Appli
c	10	17	12.4	92	1	US-08-239-427A-1	Sequence 1, Appli
c	11	17	12.4	92	2	US-08-467-975-1	Sequence 1, Appli
c	12	17	12.4	92	3	US-08-401-355-1	Sequence 1, Appli
c	13	17	12.4	1260	3	US-09-081-689-3	Sequence 3, Appli
c	14	17	12.4	1297	4	US-08-858-207A-66	Sequence 66, Appl
c	15	17	12.4	1329	4	US-09-305-984-13	Sequence 13, Appl
c	16	17	12.4	1329	4	US-09-073-541A-13	Sequence 13, Appl
c	17	17	12.4	1329	4	US-09-493-940-13	Sequence 13, Appl
c	18	17	12.4	1526	3	US-09-081-689-1	Sequence 1, Appli
	19	17	12.4	2817	5	PCT-US93-05944-1	Sequence 1, Appli
	20	17	12.4	3356	3	US-09-379-523-4	Sequence 4, Appli
	21	17	12.4	3808	2	US-08-417-210A-79	Sequence 79, Appl
	22	17	12.4	3808	4	US-09-136-159A-79	Sequence 79, Appl
	23	17	12.4	4358	4	US-09-308-453-1	Sequence 1, Appli
c	24	17	12.4	5152	4	US-10-204-708-48	Sequence 48, Appl
	25	17	12.4	5915	4	US-10-204-708-77	Sequence 77, Appl
	26	17	12.4	7616	3	US-08-815-809-6	Sequence 6, Appli
	27	17	12.4	7797	2	US-08-816-155B-7	Sequence 7, Appli
	28	17	12.4	7797	3	US-09-079-587-7	Sequence 7, Appli
c	29	17	12.4	8900	4	US-09-305-984-25	Sequence 25, Appl
c	30	17	12.4	8900	4	US-09-073-541A-25	Sequence 25, Appl
c	31	17	12.4	8900	4	US-09-493-940-25	Sequence 25, Appl
	32	17	12.4	10962	2	US-08-816-155B-6	Sequence 6, Appli
	33	17	12.4	10962	3	US-09-079-587-6	Sequence 6, Appli
	34	17	12.4	19877	2	US-08-816-155B-8	Sequence 8, Appli
	35	17	12.4	19877	3	US-09-079-587-8	Sequence 8, Appli
	36	17	12.4	28882	4	US-08-961-527-140	Sequence 140, App
c	37	16	11.7	312	4	US-09-543-681A-1844	Sequence 1844, Ap
	38	16	11.7	402	4	US-09-621-976-113	Sequence 113, App
	39	16	11.7	439	4	US-09-397-787-294	Sequence 294, App
	40	16	11.7	661	4	US-09-702-705-219	Sequence 219, App
	41	16	11.7	661	4	US-09-736-457-219	Sequence 219, App
	42	16	11.7	661	4	US-09-614-124B-219	Sequence 219, App
	43	16	11.7	661	4	US-09-671-325-219	Sequence 219, App
	44	16	11.7	661	4	US-09-589-184-219	Sequence 219, App
	45	16	11.7	1247	4	US-09-740-035-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-484-970B-60/c
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
; NAME/KEY: unsure
; LOCATION: 11721-11761, 12294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match 85.4%; Score 117; DB 4; Length 13977;
Best Local Similarity 100.0%; Pred. No. 3.6e-51;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13306 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13247

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13246 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 13190
```

RESULT 2
US-08-836-022A-10
; Sequence 10, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Chen, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and

```
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P O Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,022A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/331,381
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GNVPN.008PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19307 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-836-022A-10

  Query Match             30.7%;  Score 42;  DB 3;  Length 19307;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          75 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 116
               ||||||||||||||||||||||||||||||||||||||||||
Db        1563 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 1604


RESULT 3
US-09-427-048A-10
; Sequence 10, Application US/09427048A
; Patent No. 6203975
;   GENERAL INFORMATION:
;       APPLICANT: Trustees of the University of Pennsylvania
;                  Wilson, James M.
;                  Fisher, Krishna J.
;                  Chen, Shu-Jen
;                  Weitzman, Matthew
;       TITLE OF INVENTION: Improved Adenovirus Virus and
;                           Methods of Use Thereof
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Cntr, P O Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/427,048A
;            FILING DATE: 21-Oct-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/836,022
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;  INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 19307 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

  Query Match             30.7%;  Score 42;  DB 3;  Length 19307;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          75 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 116
               ||||||||||||||||||||||||||||||||||||||||||
Db        1563 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 1604


RESULT 4
PCT-US93-07261-10
; Sequence 10, Application PC/TUS9307261
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John H. C. Blasdale
;      STREET:  One Giralda Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07940-1000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.5
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07261
;      FILING DATE:  19930805
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/927,531
;      FILING DATE:  07-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Blasdale, John H. C.
;      REGISTRATION NUMBER:  31,895
;      REFERENCE/DOCKET NUMBER:  DX0288K
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-822-7398
;      TELEFAX:  201-822-7039
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4766 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium falciparum
;      STRAIN:  Malayan Camp
;    IMMEDIATE SOURCE:
;      CLONE:  p2b1;p12-1
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..4766
PCT-US93-07261-10

  Query Match             13.9%;  Score 19;  DB 5;  Length 4766;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 TAAAGGAAAAAGAAAATAA 23
              |||||||||||||||||||
Db        190 TAAAGGAAAAAGAAAATAA 208


RESULT 5
US-09-107-532A-866
; Sequence 866, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 866:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 750 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...750
;        SEQUENCE DESCRIPTION: SEQ ID NO: 866:
US-09-107-532A-866

  Query Match             13.1%;  Score 18;  DB 4;  Length 750;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 CAATGGACAAGTGGTGAA 43
              ||||||||||||||||||
Db        117 CAATGGACAAGTGGTGAA 134


RESULT 6
US-08-935-886-7
; Sequence 7, Application US/08935886
; Patent No. 5994625
;  GENERAL INFORMATION:
;    APPLICANT:  Melchers, Leo S.
;    APPLICANT:  Sela-Buurlage, Marianne B.
;    APPLICANT:  Bres-Vloemans, Alexandra A.
;    APPLICANT:  Ponstein, Anne S.
;    APPLICANT:  Apotheker-de Groot, Marion
;    APPLICANT:  Cornelissen, Bernardus J. C.
;    TITLE OF INVENTION:  Antifungal Chitin Binding Proteins and DNA
;    TITLE OF INVENTION:  Coding Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ladas & Parry
;      STREET:  26 West 61st Street
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10023
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch,  14.4 MB storage
;      COMPUTER:  IBM PC/XT/AT or compatibles
;      OPERATING SYSTEM:  DOS  6.20
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/935,886
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/411,640
;      FILING DATE:  05-APR-1995
;      APPLICATION NUMBER:  PCT/EP93/02790
;      FILING DATE:  05-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mass, Clifford J.
;      REGISTRATION NUMBER:  30,086
;      REFERENCE/DOCKET NUMBER:  U-010139-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 708-1800
;      TELEFAX:  (212) 246-8959
;      TELEX:  233288
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  889 Base Pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nicotiana tabacum
;      STRAIN:  Samsun NN
;      DEVELOPMENTAL STAGE:  Mature
;      TISSUE TYPE:  Leaf, wounded
;    IMMEDIATE SOURCE:
;      LIBRARY:  lambda zap
;      CLONE:  CBP4.4
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  16..639
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..14
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  866..889
;      OTHER INFORMATION:  /function= "XhoI-linker"
```

```
US-08-935-886-7

  Query Match             13.1%;  Score 18;  DB 2;  Length 889;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 AAAGGAAAAAGAAAATAA 23
              ||||||||||||||||||
Db        707 AAAGGAAAAAGAAAATAA 724


RESULT 7
US-08-935-886-13
; Sequence 13, Application US/08935886
; Patent No. 5994625
;  GENERAL INFORMATION:
;    APPLICANT:  Melchers, Leo S.
;    APPLICANT:  Sela-Buurlage, Marianne B.
;    APPLICANT:  Bres-Vloemans, Alexandra A.
;    APPLICANT:  Ponstein, Anne S.
;    APPLICANT:  Apotheker-de Groot, Marion
;    APPLICANT:  Cornelissen, Bernardus J. C.
;    TITLE OF INVENTION:  Antifungal Chitin Binding Proteins and DNA
;    TITLE OF INVENTION:  Coding Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ladas & Parry
;      STREET:  26 West 61st Street
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10023
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch,  14.4 MB storage
;      COMPUTER:  IBM PC/XT/AT or compatibles
;      OPERATING SYSTEM:  DOS  6.20
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/935,886
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/411,640
;      FILING DATE:  05-APR-1995
;      APPLICATION NUMBER:  PCT/EP93/02790
;      FILING DATE:  05-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mass, Clifford J.
;      REGISTRATION NUMBER:  30,086
;      REFERENCE/DOCKET NUMBER:  U-010139-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 708-1800
;      TELEFAX:  (212) 246-8959
;      TELEX:  233288
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  889 Base Pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nicotiana tabacum
;      STRAIN:  Samsun NN
;      TISSUE TYPE:  Leaf, wounded
;    IMMEDIATE SOURCE:
;      CLONE:  CBP4.4T
US-08-935-886-13

  Query Match             13.1%;  Score 18;  DB 2;  Length 889;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 AAAGGAAAAAGAAAATAA 23
              ||||||||||||||||||
Db        726 AAAGGAAAAAGAAAATAA 743


RESULT 8
US-09-134-001C-1838
; Sequence 1838, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1838
;   LENGTH: 2076
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1838

  Query Match             13.1%;  Score 18;  DB 4;  Length 2076;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TTATAAAGGAAAAAGAAA 19
              ||||||||||||||||||
Db       1374 TTATAAAGGAAAAAGAAA 1391


RESULT 9
US-07-839-751-1/c
; Sequence 1, Application US/07839751
; Patent No. 5439809
;  GENERAL INFORMATION:
;    APPLICANT:  Haynes, Joel
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Production of Genetically-Engineered
;    TITLE OF INVENTION:  Vaccines for Aids and Retroviral Diseases
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/839,751
;      FILING DATE:  19920615
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-267 MIS:jc
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-839-751-1

  Query Match             12.4%;  Score 17;  DB 1;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 AGGAAAAAGAAAATAAC 24
              |||||||||||||||||
Db         73 AGGAAAAAGAAAATAAC 57


RESULT 10
US-08-239-427A-1/c
; Sequence 1, Application US/08239427A
; Patent No. 5571712
;  GENERAL INFORMATION:
;    APPLICANT:  Haynes, Joel
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Rovinski, Benjamin
;    APPLICANT:  Cao, Shi Xian
;    TITLE OF INVENTION:  Production of Genetically-Engineered
;    TITLE OF INVENTION:  Vaccines for Aids and Retroviral Diseases
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/239,427A
;      FILING DATE:  06-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 839,751
;      FILING DATE:  12-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-332 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-239-427A-1

  Query Match             12.4%;  Score 17;  DB 1;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 AGGAAAAAGAAAATAAC 24
              |||||||||||||||||
Db         73 AGGAAAAAGAAAATAAC 57


RESULT 11
US-08-467-975-1/c
; Sequence 1, Application US/08467975
; Patent No. 5985641
;  GENERAL INFORMATION:
;    APPLICANT:  Haynes, Joel
;    APPLICANT:  Klein, Michel H
;    APPLICANT:  Rovinski, Benjamin
;    APPLICANT:  Cao, Shi Xian
;    TITLE OF INVENTION:  Production of Genetically-Engineered
;    TITLE OF INVENTION:  Vaccines for Aids and Other Retroviral Diseases
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,975
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,427
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/839,751
;      FILING DATE:  12-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8912123.7
;      FILING DATE:  13-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I.
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-456 MIS:vg
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-467-975-1

  Query Match             12.4%;  Score 17;  DB 2;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 AGGAAAAAGAAAATAAC 24
              |||||||||||||||||
Db         73 AGGAAAAAGAAAATAAC 57


RESULT 12
US-08-401-355-1/c
; Sequence 1, Application US/08401355
; Patent No. 6291227
;  GENERAL INFORMATION:
;    APPLICANT:  Haynes, Joel
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Production of Genetically-Engineered
;    TITLE OF INVENTION:  Vaccines for Aids and Retroviral Diseases
;    NUMBER OF SEQUENCES:  2
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5R 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,355
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/839,751
;      FILING DATE:  15-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-424 MIS:as
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;      TELEX:  065-24567 SIMBAS
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  92 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-401-355-1

  Query Match             12.4%;  Score 17;  DB 3;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 AGGAAAAAGAAAATAAC 24
              |||||||||||||||||
Db         73 AGGAAAAAGAAAATAAC 57


RESULT 13
US-09-081-689-3/c
; Sequence 3, Application US/09081689
; Patent No. 6165992
;  GENERAL INFORMATION:
;    APPLICANT:  Wallis, Nicola G.
;    APPLICANT:  Zalacain, Madgalenaa
;    APPLICANT:  Throup, John
;    APPLICANT:  Biswas, Sanjoy
;    TITLE OF INVENTION:  Histidine Kinase
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert, Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2793
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/081,689
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/048,347
;      FILING DATE:  30-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Falk, Stephen T
;      REGISTRATION NUMBER:  36,795
;      REFERENCE/DOCKET NUMBER:  GM10009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2488
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1260 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-081-689-3

  Query Match             12.4%;  Score 17;  DB 3;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 AAGGAAAAAGAAAATAA 23
              |||||||||||||||||
Db        631 AAGGAAAAAGAAAATAA 615


RESULT 14
US-08-858-207A-66/c
; Sequence 66, Application US/08858207A
; Patent No. 6348328
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Stodola, Robert
;    TITLE OF INVENTION:  No. 6348328el Compounds
;    NUMBER OF SEQUENCES:  552
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,207A
;      FILING DATE:  09-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017670
;      FILING DATE:  14-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50475
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1297 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-858-207A-66
```

Query Match 12.4%; Score 17; DB 4; Length 1297;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          7 AAGGAAAAAGAAAATAA 23
              |||||||||||||||||
Db        631 AAGGAAAAAGAAAATAA 615
```

RESULT 15
US-09-305-984-13/c
; Sequence 13, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-305-984-13

Query Match 12.4%; Score 17; DB 4; Length 1329;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          7 AAGGAAAAAGAAAATAA 23
              |||||||||||||||||
Db        103 AAGGAAAAAGAAAATAA 87
```

Search completed: April 6, 2004, 17:42:27
Job time : 16.6147 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:42:46 ; Search time 64.9329 Seconds
(without alignments)
8963.144 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	137	100.0	137	6	ABK86496	Abk86496 Human Apo
c	2	137	100.0	158	6	ABK86469	Abk86469 Human apo
c	3	137	100.0	200	6	ABK86468	Abk86468 Human dys
c	4	137	100.0	455	9	ADD32504	Add32504 Human mit
	5	137	100.0	996	6	ABK86497	Abk86497 Human Apo
	6	137	100.0	1230	6	ABK86462	Abk86462 Human Apo
	7	137	100.0	1234	6	ABK86463	Abk86463 Human Apo
c	8	137	100.0	2691	6	ABK81996	Abk81996 Human dys
c	9	137	100.0	8689	6	ABK82000	Abk82000 DNA encod
c	10	137	100.0	11443	6	ABK82002	Abk82002 DNA encod
c	11	137	100.0	13957	6	ABK81959	Abk81959 cDNA enco
c	12	137	100.0	13957	6	ABT10904	Abt10904 Human bre
c	13	137	100.0	13957	6	ABN95786	Abn95786 Gene #228
c	14	137	100.0	13957	6	ABS69900	Abs69900 Human dys
c	15	117	85.4	13977	6	ABS70403	Abs70403 Human bon
c	16	46	33.6	959	6	ABN74601	Abn74601 Bovine em
c	17	42	30.7	13815	2	AAV18885	Aav18885 Mus muscu
c	18	42	30.7	13815	6	ABK81960	Abk81960 cDNA enco
c	19	42	30.7	13815	6	ABI99799	Abi99799 Mouse isc
	20	42	30.7	19307	2	AAT27558	Aat27558 Shuttle v
	21	41	29.9	60	6	ABK86471	Abk86471 Human apo
	22	40	29.2	108	6	ABK86467	Abk86467 Human apo
c	23	22	16.1	22	6	ABK86479	Abk86479 Human apo
c	24	20	14.6	690	6	ABZ08220	Abz08220 Human leu
c	25	19	13.9	716	4	AAK64599	Aak64599 Human imm
	26	19	13.9	4766	2	AAQ70102	Aaq70102 Malarial
	27	19	13.9	7326	3	AAA70095	Aaa70095 Plasmodiu
c	28	19	13.9	42521	2	AAZ32013	Aaz32013 Human MET
c	29	19	13.9	42521	5	AAC90070	Aac90070 U64857 cD
	30	18	13.1	18	6	ABK86487	Abk86487 Human apo
c	31	18	13.1	324	2	AAZ33883	Aaz33883 Tobacco p
c	32	18	13.1	407	4	AAK75817	Aak75817 Human imm
c	33	18	13.1	408	4	AAK75816	Aak75816 Human imm
	34	18	13.1	476	9	ADB51265	Adb51265 Primary r
	35	18	13.1	492	7	ABZ76803	Abz76803 Nicotiana
c	36	18	13.1	569	3	AAC55084	Aac55084 Arabidops
	37	18	13.1	717	7	ACA33643	Aca33643 Prokaryot
	38	18	13.1	750	9	ADC91239	Adc91239 E. faeciu
c	39	18	13.1	878	6	ABN98770	Abn98770 Arabidops
	40	18	13.1	889	2	AAQ62459	Aaq62459 Chitin bi
	41	18	13.1	889	2	AAQ62455	Aaq62455 Chitin bi
c	42	18	13.1	896	3	AAC37805	Aac37805 Arabidops
c	43	18	13.1	1028	3	AAC37406	Aac37406 Arabidops
c	44	18	13.1	1155	3	AAC36681	Aac36681 Arabidops
	45	18	13.1	2076	6	ABN92375	Abn92375 Staphyloc

ALIGNMENTS

RESULT 1
ABK86496
ID ABK86496 standard; DNA; 137 BP.
XX
AC ABK86496;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human Apo-dystrophin-4 inversion sequence.
XX
KW Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3. .137
FT /*tag= a
FT /product= "Apo-dystrophin-4 peptide appearing as
FT AAU98738"
FT /partial
FT /note= "No start or stop codon shown"
FT /transl_except= (pos:21. .23,aa:Xaa)
FT /transl_except= (pos:48. .50,aa:Xaa)
FT /transl_except= (pos:93. .95,aa:Xaa)
FT /transl_except= (pos:123. .125,aa:Xaa)
FT /transl_except= (pos:129. .131,aa:Xaa)
FT /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN GB2368064-A.
XX
PD 24-APR-2002.
XX
PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
DR P-PSDB; AAU98737.
XX
PT New human regulatory polynucleotide, useful for treating disorders

PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Claim 1; Page 169; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is the apo-dystrophin-4 inversion sequence and upstream
CC genomic region
XX
SQ Sequence 137 BP; 56 A; 19 C; 32 G; 30 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 137;
Best Local Similarity 100.0%; Pred. No. 4.8e-59;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        121 TTTGATGTTAATTAATT 137
```

RESULT 2
ABK86469/c
ID ABK86469 standard; cDNA; 158 BP.
XX
AC ABK86469;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human apo-dystrophin-4 cDNA fragment with inversion breakpoint #2.
XX
KW Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia; dystrophin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_recomb 137
FT /*tag= a
FT /label= Inversion_breakpoint
XX
PN GB2368064-A.
XX
PD 24-APR-2002.
XX
PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
XX
PT New human regulatory polynucleotide, useful for treating disorders
PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Disclosure; Fig 16B; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is a apo-dystrophin-4 cDNA fragment showing an inversion
CC breakpoint (recombination signal sequence) similar to that in with human
CC dystrophin
XX
SQ Sequence 158 BP; 35 A; 35 C; 20 G; 68 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 158;
Best Local Similarity 100.0%; Pred. No. 4.8e-59;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 78

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 18

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         17 TTTGATGTTAATTAATT 1
```

RESULT 3
ABK86468/c
ID ABK86468 standard; DNA; 200 BP.
XX
AC ABK86468;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human dystrophin genomic DNA with inversion breakpoint #2.
XX
KW Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia; dystrophin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_recomb 42

```
FT                   /*tag=  a
FT                   /label= Inversion_breakpoint
FT   misc_recomb     179
FT                   /*tag=  b
FT                   /label= Inversion_breakpoint
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 16B; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is a human dystrophin genomic DNA fragment showing
CC   inversion breakpoints (recombination signal sequence) similar to that in
CC   apo-dystrophin-4
XX
SQ   Sequence 200 BP; 44 A; 40 C; 28 G; 88 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 137;  DB 6;  Length 200;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 120

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 60

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         59 TTTGATGTTAATTAATT 43


RESULT 4
ADD32504/c
ID   ADD32504 standard; DNA; 455 BP.
XX
AC   ADD32504;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human mitochondrial DNA sequence SEQ ID NO:274.
XX
KW   ds; human; array; mitochondrial; hybridisation; energy-metabolism;
KW   mitochondrial disease; oxidative phosphorylation dysfunction;
KW   oxidative stress; apoptosis; aging.
XX
OS   Homo sapiens.
XX
PN   WO2003020220-A2.
XX
PD   13-MAR-2003.
XX
PF   30-AUG-2002; 2002WO-US027886.
XX
PR   30-AUG-2001; 2001US-0316323P.
PR   31-AUG-2001; 2001CA-02356540.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wallace DC,  Levy S,  Kerstann K,  Procaccio V;
XX
DR   WPI; 2003-300821/29.
XX
PT   Array containing probes for genes involved in mitochondrial biology,
PT   useful for determining mitochondrial biology gene expression profiles for
PT   use in diagnosing pathologies and identifying biochemical pathways.
XX
PS   Claim 2; SEQ ID NO 274; 201pp; English.
XX
CC   The invention relates to a novel array comprising at least two isolated
CC   nucleotide molecules, each molecule having a sequence capable of uniquely
CC   hybridising to a nucleic acid molecule which is an expression product of
CC   a gene involved in mitochondrial biology. The array comprises two or more
CC   isolated nucleic acid molecules or spots, each molecule having a sequence
CC   chosen from sequence of 994 human probes and 2046 mouse probes. An array
CC   of the invention is useful for determining an expression profile of a
CC   mouse or human sample containing nucleic acid, by contacting the array
CC   with the sample under conditions allowing selective hybridisation, and
CC   measuring hybridisation of nucleic acid in the sample to the array to
CC   produce an expression profile. The array is also useful for determining
CC   an expression profile of a first labelled sample containing nucleic acid
CC   relative to a second, differently labelled sample containing nucleic
CC   acid. The second sample is a reference or a standard. An array is useful
CC   for determining an expression profile diagnostic of an energy-metabolism-
CC   related physiological condition. An array of the invention is useful for
CC   determining mitochondrial biology gene expression profiles of organisms,
CC   such as human, mice and closely related species, tissue and organs of
CC   such organisms, which are useful for determining expression profiles
CC   diagnostic of energy metabolism-related physiological conditions,
CC   diagnosing such physiological conditions, identifying biochemical
CC   pathways, genes, and mutations involved in such physiological conditions,
CC   identifying therapeutic agents useful for preventing and/or treating such
CC   physiological conditions, evaluating and/or monitoring the efficacy of
CC   such therapies, and creating and identifying animal models of human
CC   energy metabolism-related physiological conditions. An array is also
CC   useful for defining expression signatures or profiles for mitochondrial
CC   diseases, as well as distinguishing clinical disorders that result from
CC   oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC   apoptosis and aging. An array of the invention contains probes of genes
CC   not previously recognised to participate in mitochondrial biology. The
CC   sequences shown in ADD32231-ADD33223 represent human mitochondrial DNA
CC   clones used to make the probes of the invention. Some sequences are not
CC   present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC   1906, 2408 and 2643.
XX
SQ   Sequence 455 BP; 127 A; 80 C; 63 G; 185 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 137;  DB 9;  Length 455;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-59;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 96

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 36

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         35 TTTGATGTTAATTAATT 19


RESULT 5
ABK86497
ID   ABK86497 standard; cDNA; 996 BP.
XX
AC   ABK86497;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 cDNA.
XX
KW   Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .996
FT                   /*tag=  a
FT                   /product= "Apo-dystrophin-4 protein appearing as
FT                   AAU98739"
FT                   /partial
FT                   /note= "No start or stop codon shown"
FT                   /transl_except= (pos:7. .9,aa:Xaa)
FT                   /transl_except= (pos:49. .51,aa:Xaa)
FT                   /transl_except= (pos:187. .189,aa:Xaa)
FT                   /transl_except= (pos:202. .204,aa:Xaa)
FT                   /transl_except= (pos:217. .219,aa:Xaa)
FT                   /transl_except= (pos:253. .255,aa:Xaa)
FT                   /transl_except= (pos:265. .270,aa:Xaa-Xaa)
FT                   /transl_except= (pos:394. .396,aa:Xaa)
FT                   /transl_except= (pos:547. .549,aa:Xaa)
FT                   /transl_except= (pos:565. .567,aa:Xaa)
FT                   /transl_except= (pos:616. .618,aa:Xaa)
FT                   /transl_except= (pos:649. .651,aa:Xaa)
FT                   /transl_except= (pos:682. .684,aa:Xaa)
FT                   /transl_except= (pos:709. .711,aa:Xaa)
FT                   /transl_except= (pos:790. .792,aa:Xaa)
FT                   /transl_except= (pos:841. .843,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:880. .882,aa:Xaa)
FT                   /transl_except= (pos:907. .909,aa:Xaa)
FT                   /transl_except= (pos:952. .954,aa:Xaa)
FT                   /transl_except= (pos:982. .984,aa:Xaa)
FT                   /transl_except= (pos:988. .990,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98739.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Claim 5; Page 170-172; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the cDNA sequence for human apo-dystrophin-4
CC   containing a plurality of stop codons some of which may be read through
CC   due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 137;  DB 6;  Length 996;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 919

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 979

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        980 TTTGATGTTAATTAATT 996


RESULT 6
ABK86462
ID   ABK86462 standard; cDNA; 1230 BP.
XX
AC   ABK86462;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 full length cDNA.
XX
KW   Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1230
FT                   /*tag=  a
FT                   /product= "Apo-dystrophin-4"
FT                   /partial
FT                   /note= "No start or stop codon shown"
```

```
FT                   /transl_except= (pos:142. .144,aa:Xaa)
FT                   /transl_except= (pos:241. .244,aa:Xaa)
FT                   /transl_except= (pos:282. .284,aa:Xaa)
FT                   /transl_except= (pos:421. .423,aa:Xaa)
FT                   /transl_except= (pos:436. .438,aa:Xaa)
FT                   /transl_except= (pos:451. .453,aa:Xaa)
FT                   /transl_except= (pos:487. .489,aa:Xaa)
FT                   /transl_except= (pos:499. .504,aa:Xaa-Xaa)
FT                   /transl_except= (pos:628. .630,aa:Xaa)
FT                   /transl_except= (pos:781. .783,aa:Xaa)
FT                   /transl_except= (pos:799. .801,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:883. .885,aa:Xaa)
FT                   /transl_except= (pos:916. .920,aa:Xaa)
FT                   /transl_except= (pos:945. .947,aa:Xaa)
FT                   /transl_except= (pos:1024. .1026,aa:Xaa)
FT                   /transl_except= (pos:1075. .1077,aa:Xaa)
FT                   /transl_except= (pos:1084. .1086,aa:Xaa)
FT                   /transl_except= (pos:1114. .1116,aa:Xaa)
FT                   /transl_except= (pos:1141. .1143,aa:Xaa)
FT                   /transl_except= (pos:1186. .1188,aa:Xaa)
FT                   /transl_except= (pos:1216. .1218,aa:Xaa)
FT                   /transl_except= (pos:1222. .1224,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98729.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 4; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the full length cDNA sequence for human apo-
CC   dystrophin-4 containing a plurality of stop codons some of which may be
CC   read through due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 1230 BP; 404 A; 189 C; 259 G; 378 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 137;  DB 6;  Length 1230;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1094 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1153

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1154 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1213

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1214 TTTGATGTTAATTAATT 1230


RESULT 7
ABK86463
ID   ABK86463 standard; cDNA; 1234 BP.
XX
AC   ABK86463;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 full cDNA fragment #1.
XX
KW   Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 9; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is a human apo-dystrophin-4 fragment matching a region
CC   of mouse genomic DNA in the Apo-dystrophin-4 gene region
```

```
XX
SQ   Sequence 1234 BP; 405 A; 190 C; 260 G; 379 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 1234;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1157

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1217

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1218 TTTGATGTTAATTAATT 1234


RESULT 8
ABK81996/c
ID   ABK81996 standard; DNA; 2691 BP.
XX
AC   ABK81996;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Human dystrophin 3' untranslated region.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Disclosure; Fig 10; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a human dystrophin polynucleotide
CC   sequence used in the creation of the mini-dystrophin peptides of the
CC   invention
XX
SQ   Sequence 2691 BP; 860 A; 448 C; 461 G; 922 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 2691;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 9
ABK82000/c
ID   ABK82000 standard; DNA; 8689 BP.
XX
AC   ABK82000;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding mini-dystrophin protein deltaH2-R19.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Disclosure; Fig 15; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a mini-dystrophin sequence of the
CC   invention
XX
SQ   Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 137;  DB 6;  Length 8689;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 7972

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 7912
```

```
Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       7911 TTTGATGTTAATTAATT 7895


RESULT 10
ABK82002/c
ID   ABK82002 standard; DNA; 11443 BP.
XX
AC   ABK82002;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding mini-dystrophin protein deltaR9-R16.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Disclosure; Fig 17; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a mini-dystrophin sequence of the
CC   invention
XX
SQ   Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10785 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 10726

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10725 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 10666

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      10665 TTTGATGTTAATTAATT 10649


RESULT 11
ABK81959/c
ID   ABK81959 standard; DNA; 13957 BP.
XX
AC   ABK81959;
XX
DT   13-AUG-2002  (first entry)
XX
DE   cDNA encoding human dystrophin.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Example 2; Fig 1; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a human dystrophin polynucleotide
CC   sequence used in the creation of the mini-dystrophin peptides of the
CC   invention
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 12
ABT10904/c
ID   ABT10904 standard; cDNA; 13957 BP.
XX
AC   ABT10904;
XX
DT   04-DEC-2002  (first entry)
XX
```

```
DE   Human breast cancer associated coding sequence SEQ ID NO: 1038.
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 1038; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 13
ABN95786/c
ID   ABN95786 standard; DNA; 13957 BP.
XX
AC   ABN95786;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2284 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US030589.
XX
PR   02-OCT-2000; 2000US-0237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer, hepatocellular
PT   carcinoma or metastatic liver tumor in a patient, involves detecting the
PT   level of expression of two or more genes in a liver tissue sample.
XX
PS   Claim 1; SEQ ID NO 2284; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-59;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 14
ABS69900/c
ID   ABS69900 standard; DNA; 13957 BP.
XX
AC   ABS69900;
XX
DT   21-NOV-2002  (first entry)
XX
DE   Human dystrophin gene.
XX
KW   Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;
```

KW adenosine deaminase deficiency; severe combined immune deficiency; PAH;
KW beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;
KW low density lipoprotein gene; familial hypercholesterolaemia;
KW hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;
KW phenylalanine hydroxylase gene; gene therapy; phenylketonuria;
KW dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;
KW human cystic fibrosis transmembrane conductance regulator gene;
KW antianaemic; antilipaemic; nootropic; cytostatic; dermatological; human;
KW alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;
KW lysosomal arylsulfatase A; ornithine transcarbamylase; ARSA; OTC; NP;
KW purin nucleoside phosphorylase; gene; ds.
XX
OS Homo sapiens.
XX
PN US2002102731-A1.
XX
PD 01-AUG-2002.
XX
PF 12-FEB-2001; 2001US-00782378.
XX
PR 02-OCT-2000; 2000US-0237747P.
XX
PA (UYNY) UNIV NEW YORK STATE RES FOUND.
XX
PI Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;
XX
DR WPI; 2002-690619/74.
XX
PT Producing vector, by introducing vector having nucleotide sequence,
PT adenovirus inverted terminal repeats and packaging sequence, and adeno-
PT associated virus terminal repeat, into cell, and culturing cell.
XX
PS Disclosure; Page 122-128; 191pp; English.
XX
CC The present invention relates to a new method of producing a vector. The
CC method involves introducing recombinant vector having nucleotide sequence
CC (NS) having 5' and 3' end, left and right inverted terminal repeats of
CC adenovirus flanking NS, adenovirus packaging sequence linked to inverted
CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'
CC end of NS, into cell expressing adenovirus early gene lacking from vector
CC ; and culturing cell to produce another vector. The method is useful for
CC generating vectors, especially mAd vectors. The method is useful in
CC transferring nucleotide sequences of interest into a cell, for gene
CC transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.
CC The nucleotide sequences are useful for treating diseases associated with
CC it, i.e. adenosine deaminase gene associated with adenosine deaminase
CC deficiency with severe combined immune deficiency, beta-chain of
CC haemoglobin gene associated with beta-thalassaemia and sickle cell
CC disease, receptor for low density lipoprotein gene associated with
CC familial hypercholesterolaemia, hypoxanthine-guanine
CC phosphoribosyltransferase associated with Lesch-Nyhan syndrome,
CC phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,
CC dystrophin gene associated with muscular dystrophy, and human cystic
CC fibrosis transmembrane conductance regulator gene associated with cystic
CC fibrosis. The present nucleic acid sequence represents a human disease
CC gene sequence that was used in the methods of the invention
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 13957;
Best Local Similarity 100.0%; Pred. No. 4.6e-59;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy 121 TTTGATGTTAATTAATT 137
|||||||||||||||||
Db 13179 TTTGATGTTAATTAATT 13163

RESULT 15
ABS70403/c
ID ABS70403 standard; cDNA; 13977 BP.
XX
AC ABS70403;
XX
DT 27-NOV-2002 (first entry)
XX
DE Human bone remodelling gene #60.
XX
KW Bone remodelling; osteoporosis; human; gene; ss.
XX
OS Homo sapiens.
XX
PN US6426186-B1.
XX
PD 30-JUL-2002.
XX
PF 18-JAN-2000; 2000US-00484970.
XX
PR 18-JAN-2000; 2000US-00484970.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Volkmuth W, Walker MG;
XX
DR WPI; 2002-673014/72.
XX
PT A combination of polynucleotides which are co-expressed with genes known
PT to be involved in bone remodeling and osteoporosis are useful in an array
PT for the diagnosis of bone remodeling and osteoporosis associated
PT disorders.
XX
PS Claim 1; Col 185-198; 206pp; English.
XX
CC The invention relates to a combination comprising a number of
CC substantially purified and isolated polynucleotides which are co-
CC expressed with genes known to be involved in bone remodelling and
CC osteoporosis. The invention is used to diagnose disorders associated with
CC bone remodelling or osteoporosis. ABS70344-ABS70512 represent human bone
CC remodelling genes of the invention
XX
SQ Sequence 13977 BP; 4596 A; 2765 C; 3120 G; 3453 T; 0 U; 43 Other;

Query Match 85.4%; Score 117; DB 6; Length 13977;
Best Local Similarity 100.0%; Pred. No. 5.3e-49;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13306 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13247

Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 117
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13246 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 13190

Search completed: April 6, 2004, 15:10:51
Job time : 66.9329 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:12:12 ; Search time 505.195 Seconds
(without alignments)
8098.097 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	137	100.0	373	10	AW580404	AW580404 PM2-HT045
c	2	137	100.0	396	14	H89576	H89576 yw28c12.r1
c	3	137	100.0	455	9	AA427831	AA427831 zw49d12.r
c	4	137	100.0	462	14	CB045405	CB045405 NISC_gc10
c	5	137	100.0	473	14	CA389721	CA389721 cs101f03.
c	6	137	100.0	489	28	AQ015185	AQ015185 CIT-HSP-2
c	7	137	100.0	647	10	AW664684	AW664684 hi84e10.x
c	8	137	100.0	696	10	AW950480	AW950480 EST362550
c	9	137	100.0	710	9	AV725574	AV725574 AV725574
c	10	137	100.0	727	12	BG567176	BG567176 602589640
c	11	137	100.0	740	14	CD356811	CD356811 AGENCOURT
c	12	137	100.0	745	14	CD357556	CD357556 AGENCOURT
c	13	137	100.0	1121	12	BM546012	BM546012 AGENCOURT
c	14	136	99.3	743	14	CD110642	CD110642 AGENCOURT
	15	120	87.6	352	10	AW385154	AW385154 PM2-HT045
c	16	112	81.8	862	14	CB962272	CB962272 AGENCOURT
	17	111	81.0	353	10	AW580423	AW580423 PM2-HT045
c	18	88	64.2	911	12	BI752714	BI752714 603028327
c	19	86	62.8	375	10	AW607064	AW607064 PM2-HT045
c	20	57	41.6	551	9	AA621932	AA621932 nq20b10.s
c	21	54	39.4	270	14	F25920	F25920 HSPD13162 H
c	22	54	39.4	841	10	BF791019	BF791019 602251072
c	23	48	35.0	265	14	F18701	F18701 HSPD03293 H
c	24	46	33.6	631	14	CB465228	CB465228 726497 MA
c	25	43	31.4	641	9	AI528613	AI528613 mq87c07.x
c	26	42	30.7	268	10	BB178429	BB178429 BB178429
c	27	42	30.7	291	10	BB174122	BB174122 BB174122
c	28	42	30.7	291	10	BB339107	BB339107 BB339107
c	29	42	30.7	296	10	BB041920	BB041920 BB041920
c	30	42	30.7	329	10	BB085328	BB085328 BB085328
	31	42	30.7	390	9	AI324317	AI324317 mq87c07.y
c	32	42	30.7	607	10	BE370292	BE370292 601222549
	33	42	30.7	610	9	AA146038	AA146038 mq87c07.r
c	34	42	30.7	643	14	CB850037	CB850037 MRA-1794
c	35	42	30.7	735	14	CB570707	CB570707 AGENCOURT
c	36	42	30.7	4437	11	AK036936	AK036936 Mus muscu
c	37	40	29.2	734	10	BF674672	BF674672 602136802
	38	40	29.2	924	13	BX455062	BX455062 BX455062
c	39	32	23.4	210	10	BB171938	BB171938 BB171938
c	40	32	23.4	418	13	BY377466	BY377466 BY377466
c	41	31	22.6	293	10	BB307434	BB307434 BB307434
c	42	30	21.9	651	13	BU338697	BU338697 603515030
c	43	30	21.9	733	13	BU257193	BU257193 603744569
c	44	27	19.7	945	10	BF180441	BF180441 601805231
c	45	25	18.2	306	10	BB086855	BB086855 BB086855

ALIGNMENTS

RESULT 1
AW580404/c
LOCUS AW580404 373 bp mRNA linear EST 16-MAR-2000
DEFINITION PM2-HT0451-080100-003-h09 HT0451 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW580404
VERSION AW580404.1 GI:7255453
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 373)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM2&t2=PM2-HT0451-080100-003-h09&t3=2000-01-08&t4=1)

Seq primer: puc 18 forward
High quality sequence stop: 373.

FEATURES Location/Qualifiers
source 1. .373
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0451"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 100.0%; Score 137; DB 10; Length 373;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 310

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 250

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        249 TTTGATGTTAATTAATT 233
```

RESULT 2
H89576/c
LOCUS H89576 396 bp mRNA linear EST 28-NOV-1995
DEFINITION yw28c12.r1 Morton Fetal Cochlea Homo sapiens cDNA clone IMAGE:253558 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.
ACCESSION H89576
VERSION H89576.1 GI:1079922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 300
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1088 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 300.

FEATURES Location/Qualifiers
source 1. .396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3891011"
/db_xref="taxon:9606"
/clone="IMAGE:253558"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned unidirectionally. Primer: Oligo dT. Fetal cochlea, normal. 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP XR Vector. Library constructed by N. Robertson, C. Morton. ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 396;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 205

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 145

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        144 TTTGATGTTAATTAATT 128
```

RESULT 3
AA427831/c
LOCUS AA427831 455 bp mRNA linear EST 16-OCT-1997
DEFINITION zw49d12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:773399 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.
ACCESSION AA427831
VERSION AA427831.1 GI:2111628
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 391.

FEATURES Location/Qualifiers
source 1. .455
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:773399"

/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 100.0%; Score 137; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 96

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 36

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         35 TTTGATGTTAATTAATT 19
```

RESULT 4
CB045405/c
LOCUS CB045405 462 bp mRNA linear EST 17-JAN-2003
DEFINITION NISC_gc10c01.y1 NCI_CGAP_Co17 Homo sapiens cDNA clone IMAGE:3218281 5', mRNA sequence.
ACCESSION CB045405
VERSION CB045405.1 GI:27783692
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 462)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM7870 row: E column: 2
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1. .462
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3218281"
/tissue_type="juvenile granulosa tumor"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Co17"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 462;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 373

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 313

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        312 TTTGATGTTAATTAATT 296
```

RESULT 5
CA389721/c
LOCUS CA389721 473 bp mRNA linear EST 06-NOV-2002
DEFINITION cs101f03.y3 Human Retinal pigment epithelium/choroid cDNA (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs101f03 5', mRNA sequence.
ACCESSION CA389721
VERSION CA389721.1 GI:24720152
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A., Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: Over 6000 non-redundant transcripts, novel genes and splice variants
JOURNAL Mol. Vis. 8 (4), 205-220 (2002)
MEDLINE 22103460
PUBMED 12107410
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 101 row: f column: 03
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1. .473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="cs101f03"
/tissue_type="RPE/choroid"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA (Un-normalized, unamplified): cs"
/note="Organ: Eye; Vector: pCMVSPORT6; Two different donor eyes (75-80 years old) yielded approximately 600 mg of dissected RPE/choroid tissue. This in turn yielded 340 ug of total RNA and 7 ug of mRNA. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies (Rockville, MD; now part of Invitrogen Corp), essentially following the protocols of the SuperScript Plasmid System (Invitrogen Corp. <http://www.invitrogen.com/>). The library code designation was cs. For this library, cDNA inserts were cloned into the NotI/MluI sites of the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

```
ORIGIN

  Query Match             100.0%;  Score 137;  DB 14;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 172

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 112

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        111 TTTGATGTTAATTAATT 95


RESULT 6
AQ015185/c
LOCUS       AQ015185                489 bp    DNA     linear   GSS 09-JUN-1998
DEFINITION  CIT-HSP-2310F4.TR CIT-HSP Homo sapiens genomic clone 2310F4,
            genomic survey sequence.
ACCESSION   AQ015185
VERSION     AQ015185.1  GI:3193921
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2310F4.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2310F4"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 28;  Length 489;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 299

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 239

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        238 TTTGATGTTAATTAATT 222


RESULT 7
AW664684/c
LOCUS       AW664684                647 bp    mRNA    linear   EST 06-APR-2000
DEFINITION  hi84e10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2979018 3' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA
            sequence.
ACCESSION   AW664684
VERSION     AW664684.1  GI:7457227
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2979018"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             100.0%;  Score 137;  DB 10;  Length 647;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 92

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 32

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db         31 TTTGATGTTAATTAATT 15


RESULT 8
AW950480/c
```

```
LOCUS       AW950480                696 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST362550 MAGE resequences, MAGA Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW950480
VERSION     AW950480.1  GI:8140134
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 15
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGA"
                     /note="Vector: pBluescriptSKm"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 10;  Length 696;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 513

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 453

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        452 TTTGATGTTAATTAATT 436


RESULT 9
AV725574/c
LOCUS       AV725574                710 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV725574 HTC Homo sapiens cDNA clone HTCBTD10 5', mRNA sequence.
ACCESSION   AV725574
VERSION     AV725574.1  GI:10831099
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HTCBTD10"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="HTC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 9;  Length 710;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 337

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 277

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        276 TTTGATGTTAATTAATT 260


RESULT 10
BG567176/c
LOCUS       BG567176                727 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602589640F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723833 5',
            mRNA sequence.
ACCESSION   BG567176
VERSION     BG567176.1  GI:13574829
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1583  row: p  column: 10
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4723833"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
```

5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

```
  Query Match             100.0%;  Score 137;  DB 12;  Length 727;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 278

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 218

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        217 TTTGATGTTAATTAATT 201
```

RESULT 11
CD356811/c

```
LOCUS       CD356811                740 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  AGENCOURT_14253503 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30404795 5', mRNA sequence.
ACCESSION   CD356811
VERSION     CD356811.1  GI:31128222
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM190  row: o  column: 12
            High quality sequence stop: 578.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30404795"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: Blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.5-4.0 kb). 14/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match             100.0%;  Score 137;  DB 14;  Length 740;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 491

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 431

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        430 TTTGATGTTAATTAATT 414
```

RESULT 12
CD357556/c

```
LOCUS       CD357556                745 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  AGENCOURT_14253482 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30402369 5', mRNA sequence.
ACCESSION   CD357556
VERSION     CD357556.1  GI:31128967
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM184  row: j  column: 10
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30402369"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: Blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.5-4.0 kb). 14/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match             100.0%;  Score 137;  DB 14;  Length 745;
  Best Local Similarity   100.0%;  Pred. No. 3e-57;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 201

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 141

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        140 TTTGATGTTAATTAATT 124


RESULT 13
BM546012/c
LOCUS       BM546012                1121 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6505286 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5588419
            5', mRNA sequence.
ACCESSION   BM546012
VERSION     BM546012.1  GI:18778623
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12359  row: h  column: 20
            High quality sequence start: 50
            High quality sequence stop: 764.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5588419"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
ORIGIN

  Query Match             100.0%;  Score 137;  DB 12;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-57;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 455

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 395

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db        394 TTTGATGTTAATTAATT 378


RESULT 14
CD110642/c
LOCUS       CD110642                743 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  AGENCOURT_13995397 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30373580 5', mRNA sequence.
ACCESSION   CD110642
VERSION     CD110642.1  GI:30754851
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM167  row: j  column: 21
            High quality sequence stop: 578.
FEATURES             Location/Qualifiers
     source          1. .743
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30373580"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.5-4.0 kb). 14/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             99.3%;  Score 136;  DB 14;  Length 743;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-57;
  Matches  136;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 79

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 19

Qy        121 TTTGATGTTAATTAAT 136
              ||||||||||||||||
Db         18 TTTGATGTTAATTAAT 3


RESULT 15
AW385154
```

```
LOCUS       AW385154                352 bp    mRNA    linear   EST 04-FEB-2000
DEFINITION  PM2-HT0451-281299-001-a07 HT0451 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW385154
VERSION     AW385154.1  GI:6889813
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM2&t2=PM2-HT0451-
            281299-001-a07&t3=1999-12-28&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 352.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0451"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN
```

Query Match 87.6%; Score 120; DB 10; Length 352;
Best Local Similarity 100.0%; Pred. No. 7.9e-49;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         18 AAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGGAAC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGGAAC 60

Qy         78 ACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 120
```

Search completed: April 6, 2004, 17:40:07
Job time : 506.195 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 13:02:17 ; Search time 574.481 Seconds
(without alignments)
10336.278 Million cell updates/sec

Title: US-09-966-264D-1
Perfect score: 137
Sequence: 1 attataaaggaaaaagaaaa.....gtgtttgatgttaattaatt 137

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	137	100.0	2148	11	G15848	G15848 human STS C
c	2	137	100.0	2563	9	AK129855	AK129855 Homo sapi
c	3	137	100.0	2691	6	AX538618	AX538618 Sequence
c	4	137	100.0	4658	9	BC028720	BC028720 Homo sapi
c	5	137	100.0	8689	6	AX538622	AX538622 Sequence
c	6	137	100.0	11443	6	AX538624	AX538624 Sequence
c	7	137	100.0	13957	6	AX409637	AX409637 Sequence
c	8	137	100.0	13957	6	AX538581	AX538581 Sequence
c	9	137	100.0	13957	9	HUMDYS	M18533 Homo sapien
	10	137	100.0	98056	9	AC006061	AC006061 Homo sapi
c	11	117	85.4	13977	6	AR220819	AR220819 Sequence
	12	65	47.4	212120	2	AC023414	AC023414 Homo sapi
c	13	46	33.6	13887	4	AF070485	AF070485 Canis fam
c	14	46	33.6	272578	2	AC108338	AC108338 Rattus no
	15	46	33.6	279539	2	AC114184	AC114184 Rattus no
c	16	42	30.7	13815	6	AX306153	AX306153 Sequence
c	17	42	30.7	13815	6	AX538582	AX538582 Sequence
c	18	42	30.7	13815	10	MUSDYSA	M68859 Mouse dystr
	19	42	30.7	19307	6	AR093392	AR093392 Sequence
	20	42	30.7	19307	6	AR142592	AR142592 Sequence
c	21	42	30.7	189131	10	AL645477	AL645477 Mouse DNA
c	22	30	21.9	13575	5	GGDYS	X13369 Chicken mRN
	23	22	16.1	208367	2	BX682542	BX682542 Mus muscu
	24	21	15.3	157446	2	AC118861	AC118861 Rattus no
c	25	20	14.6	89871	9	AL603824	AL603824 Human DNA
c	26	20	14.6	94605	9	AF410480S2	AF360695 Homo sapi
c	27	20	14.6	114226	2	AC004710	AC004710 Plasmodiu
c	28	20	14.6	126467	9	AF487553	AF487553 Homo sapi
c	29	20	14.6	143410	9	AC087636	AC087636 Homo sapi
	30	20	14.6	145990	10	AC117837	AC117837 Mus muscu
c	31	20	14.6	157216	2	AC008244	AC008244 Homo sapi
c	32	20	14.6	159557	9	AC127898	AC127898 Homo sapi
	33	20	14.6	160750	9	HSJ177I10	AL080284 Human DNA
	34	20	14.6	161601	2	AC021147	AC021147 Homo sapi
	35	20	14.6	162612	2	AC118722	AC118722 Mus muscu
c	36	20	14.6	165660	2	AC027194	AC027194 Homo sapi
	37	20	14.6	165789	2	AL591029	AL591029 Homo sapi
c	38	20	14.6	166782	2	AC021516	AC021516 Homo sapi
c	39	20	14.6	169524	2	AC023111	AC023111 Homo sapi
	40	20	14.6	174398	9	AC009988	AC009988 Homo sapi
c	41	20	14.6	177185	9	AC016632	AC016632 Homo sapi
c	42	20	14.6	180702	2	AC121392	AC121392 Rattus no
c	43	20	14.6	181390	9	AC074237	AC074237 Homo sapi
c	44	20	14.6	182126	2	AC016639	AC016639 Homo sapi
c	45	20	14.6	187466	9	AC016322	AC016322 Homo sapi

ALIGNMENTS

RESULT 1
G15848/c
LOCUS G15848 2148 bp DNA linear STS 19-JAN-1996
DEFINITION human STS CHLC.UTR_01924_M18533.P56108 clone UTR_01924_M18533, sequence tagged site.
ACCESSION G15848
VERSION G15848.1 GI:1161737
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2148)
AUTHORS Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
TITLE Cooperative Human Linkage Center

```
  JOURNAL   Unpublished (1995)
COMMENT     Synonyms: UTR_01924_M18533, CHLC.UTR_01924_M18533.T36152
            Contact: Dr. Jeffrey C. Murray
            UofI
            The University of Iowa
            Department of Pediatrics, Iowa City, IA 52242, USA
            Tel: (319) 356-3508
            Fax: (319) 356-3347
            Email: jeff-murray@uiowa.edu

            Primer A: AACGCATTTTGGGTTGTTTA
            Primer B: GATATCACGCCAAAAGGATG
            STS size: 189
            PCR Profile:
                        denature:    30 seconds at 94 degrees C
                        annealing:   75 seconds at 55 degrees C
                        extension:   15 seconds at 72 degrees C
                        PCR cycles:  27
                        extension:    6 minutes at 72 degress C
            Protocol:
                        Template:        30ng genomic DNA
                        Primer:          each 1.5 pmole
                        dNTPs:           each 200 uM
                        Taq Polymerase:  0.3 units
                        Total Vol:       10 ul

            Buffer:
                        MgCl2:  1.5mM
                        KCl:     50mM
                        Tris:    10mM
                        pH:     8.3

                        Prepared with primer pairs derived from M18533.
FEATURES              Location/Qualifiers
     source           1. .2148
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
     STS              350. .538
     primer_bind      350. .369
     primer_bind      complement(519. .538)
ORIGIN

  Query Match             100.0%;  Score 137;  DB 11;  Length 2148;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 2
AK129855/c
LOCUS       AK129855                2563 bp    mRNA    linear   PRI 10-SEP-2003
DEFINITION  Homo sapiens cDNA FLJ26345 fis, clone HRT03668.
ACCESSION   AK129855
VERSION     AK129855.1  GI:34526478
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Suzuki,Y.,
            Hata,H., Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Nishikawa,T.,
            Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2563)
  AUTHORS   Sugano,S. and Suzuki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure, Human Genome
            Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction and 5'-end one pass sequencing: Institute of Medical
            Science, University of Tokyo, Laboratory of Genome Structure, Human
            Genome Center; 3'-end one pass sequencing: RAB; clone selection for
            full insert sequencing: RAB and Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1. .2563
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="HRT03668"
                      /tissue_type="heart"
                      /clone_lib="HRT"
                      /note="cloning vector: pME18SFL3"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 9;  Length 2563;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1905 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1846

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1845 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1786

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1785 TTTGATGTTAATTAATT 1769


RESULT 3
AX538618/c
LOCUS       AX538618                2691 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 38 from Patent WO0229056.
ACCESSION   AX538618
VERSION     AX538618.1  GI:25271161
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 38 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES              Location/Qualifiers
     source           1. .2691
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 6;  Length 2691;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-63;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 1974

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 1914

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       1913 TTTGATGTTAATTAATT 1897


RESULT 4
BC028720/c
LOCUS       BC028720                4658 bp    mRNA    linear   PRI 25-AUG-2003
DEFINITION  Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker
            types), transcript variant Dp71b, mRNA (cDNA clone IMAGE:4822807),
            complete cds.
ACCESSION   BC028720
VERSION     BC028720.1  GI:20379675
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4658)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4658)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 46 Row: j Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 5032296
            This clone has the following problem: The cds is short compared to
            the longest cds in the locus.
FEATURES             Location/Qualifiers
     source          1. .4658
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4822807"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     gene            1. .4658
                     /gene="DMD"
                     /note="synonyms: BMD, DXS142, DXS164, DXS206, DXS230,
                     DXS239, DXS268, DXS269, DXS270, DXS272"
                     /db_xref="LocusID:1756"
                     /db_xref="MIM:300377"
     CDS             130. .2037
                     /codon_start=1
                     /product="dystrophin, Dp71b isoform"
                     /protein_id="AAH28720.1"
                     /db_xref="GI:20379676"
                     /db_xref="LocusID:1756"
                     /translation="MREQLKGHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKL
                     RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVN
                     VPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGF
                     CDQRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMR
                     LEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSG
                     RVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLE
                     GDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSIS
                     PNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENR
                     NLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARM
                     QILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGS
                     QTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADDLGRAMESL
                     VSVMTDEEGAE"
     misc_feature    844. .981
                     /note="ZZ; Region: Zinc finger, ZZ type. Zinc finger
                     present in dystrophin, CBP/p300. ZZ in dystrophin binds
                     calmodulin Putative zinc finger"
                     /db_xref="CDD:pfam00569"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 9;  Length 4658;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3984 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 3925

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3924 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 3865

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       3864 TTTGATGTTAATTAATT 3848


RESULT 5
AX538622/c
LOCUS       AX538622                8689 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 42 from Patent WO0229056.
ACCESSION   AX538622
VERSION     AX538622.1  GI:25271171
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 42 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .8689
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 6;  Length 8689;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 7972

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 7912

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db       7911 TTTGATGTTAATTAATT 7895


RESULT 6
AX538624/c
LOCUS       AX538624               11443 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 44 from Patent WO0229056.
ACCESSION   AX538624
VERSION     AX538624.1  GI:25271175
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 44 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .11443
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 6;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 5e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10785 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 10726

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10725 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 10666

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      10665 TTTGATGTTAATTAATT 10649


RESULT 7
AX409637/c
LOCUS       AX409637               13957 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2284 from Patent WO0229103.
ACCESSION   AX409637
VERSION     AX409637.1  GI:21442342
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2284 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M18533"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 5e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 8
AX538581/c
LOCUS       AX538581               13957 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 1 from Patent WO0229056.
ACCESSION   AX538581
VERSION     AX538581.1  GI:25271086
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 1 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 137;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 5e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180
```

```
Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163


RESULT 9
HUMDYS/c
LOCUS       HUMDYS                13957 bp    mRNA    linear   PRI 25-MAY-2000
DEFINITION  Homo sapiens dystrophin (DMD) mRNA, complete cds.
ACCESSION   M18533 M17154 M18026 M20250
VERSION     M18533.1  GI:181856
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1699)
  AUTHORS   Koenig,M., Hoffman,E.P., Bertelson,C.J., Monaco,A.P., Feener,C. and
            Kunkel,L.M.
  TITLE     Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
            preliminary genomic organization of the DMD gene in normal and
            affected individuals
  JOURNAL   Cell 50 (3), 509-517 (1987)
  MEDLINE   87273512
   PUBMED   3607877
REFERENCE   2  (bases 1678 to 3830)
  AUTHORS   Hoffman,E.P., Monaco,A.P., Feener,C.C. and Kunkel,L.M.
  TITLE     Conservation of the Duchenne muscular dystrophy gene in mice and
            humans
  JOURNAL   Science 238 (4825), 347-350 (1987)
  MEDLINE   88018015
   PUBMED   3659917
REFERENCE   3  (bases 1 to 13957)
  AUTHORS   Koenig,M., Monaco,A.P. and Kunkel,L.M.
  TITLE     The complete sequence of dystrophin predicts a rod-shaped
            cytoskeletal protein
  JOURNAL   Cell 53 (2), 219-226 (1988)
  MEDLINE   88194521
   PUBMED   3282674
COMMENT     On May 25, 2000 this sequence version replaced gi:340693.
            Draft entry and computer-readable sequence kindly provided by
            M.Koenig, 01-APR-1988 The severity of muscular dystrophy is
            determined by the size of the deleted DNA segment. Deletions found
            in different patients were from positions 302-2200, 473-1168,
            1691-1810, and 1169-3011.
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp21.3-p21.1"
                     /tissue_type="muscle"
                     /dev_stage="fetus"
                     /note="G00-119-850"
     gene            1. .13957
                     /gene="DMD"
     CDS             209. .11266
                     /gene="DMD"
                     /codon_start=1
                     /product="dystrophin"
                     /protein_id="AAA53189.1"
                     /db_xref="GI:181857"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDV
                     DTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQ
                     QITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL
                     MESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTAH
                     QGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLM
                     DLQNQKLKELNDWLTKTEERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVN
                     SLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTE
                     EQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSLK
                     QDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVM
                     ETVTTVTTREQILVKHAQEELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEA
                     VLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSAQALVEQMVNEGVNA
                     DSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQ
                     PTTPSEPTAIKSQLKICKDEVNRLSGLQPQIERLKIQSIALKEKGQGPMFLDADFVAF
                     TNHFKQVFSDVQAREKELQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDY
                     EIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWK
                     KLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL
                     KQCRLLVSDIQTIQPSLNSVNEGGQKIKNEAEPEPASRLETELKELNTQWDHMCQQVY
                     ARKEALKGGLEKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQKAVEEMKRAK
                     EEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNGKCKTLE
                     EVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNP
                     NQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETEKS
                     LHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEA
                     AQRVLSQIDVAQKKLQDVSMKFRLFQKPANFELRLQESKMILDEVKMHLPALETKSVE
                     QEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHY
                     NELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSE
                     VAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIAVTSRAE
                     EWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELND
                     IRPKVDSTRDQAANLMANRGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELE
                     QFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQRGDNLQQRITDER
                     KREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKL
                     ASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWRE
                     IESKFAQFRRLNFAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQ
                     LLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSGRIDIIHSKKTAALQSATPVERVKL
                     QEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI
                     PENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNL
                     RWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKE
                     KLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKLKQTNLQWIKVS
                     RALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFD
                     VQETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRA
                     KQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWT
                     ELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLEELITAAQNLK
                     NKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVL
                     GQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSAD
                     DTRKVHMITENINASWRSIHKRVSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTA
                     NVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQKILRSLEGSD
                     DAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDEL
                     SRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQ
                     EPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNLHSADWQRKIDETLERLQELQEATD
                     ELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQL
                     TTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP
                     WERAISPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKA
                     LCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVD
                     MCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRL
                     GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSM
                     VWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGH
                     KMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
                     PVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESID
                     DEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEY
                     DRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDH
                     NKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSDSM
                     GEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM"
ORIGIN

  Query Match           100.0%;  Score 137;  DB 9;  Length 13957;
  Best Local Similarity  100.0%;  Pred. No. 5e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13299 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13240

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13239 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 13180

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      13179 TTTGATGTTAATTAATT 13163
```

```
RESULT 10
AC006061
LOCUS       AC006061               98056 bp    DNA     linear   PRI 01-MAY-2002
DEFINITION  Homo sapiens X BAC GSHB-19O24 (Genome Systems Human BAC Library)
            complete sequence.
ACCESSION   AC006061
VERSION     AC006061.1  GI:4204246
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98056)
  AUTHORS   Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z.,
            Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,
            Gorrell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L.,
            Kondejewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J.,
            Martinez,C., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
            Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q.,
            Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   7  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   8  (bases 1 to 98056)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 30, 1999 this sequence version replaced gi:4176317.
            INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
FEATURES             Location/Qualifiers
     source          1. .98056
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="GSHB-19O24"
     misc_feature    1. .1999
                     /note="overlaps bases 151178. .153177 of clone AC117405"
                     /function="clone overlap"
     repeat_region   40. .82
                     /rpt_family="L2"
     repeat_region   378. .507
                     /rpt_family="L1MD3"
     repeat_region   complement(518. .570)
                     /rpt_family="(GGA)n"
     repeat_region   571. .729
                     /rpt_family="L1MC4"
     repeat_region   730. .987
                     /rpt_family="MSTD"
     repeat_region   complement(991. .1400)
                     /rpt_family="L2"
     repeat_region   complement(1649. .1779)
                     /rpt_family="MIR"
     repeat_region   complement(1783. .2212)
                     /rpt_family="LTR16A"
     repeat_region   complement(3626. .3656)
                     /rpt_family="(GA)n"
     repeat_region   complement(4713. .5074)
                     /rpt_family="THE1B"
     repeat_region   5702. .7108
                     /rpt_family="L2"
     repeat_region   complement(8670. .8734)
                     /rpt_family="MLT1G"
     repeat_region   complement(9069. .9144)
                     /rpt_family="MLT1G"
     repeat_region   complement(9254. .9351)
                     /rpt_family="MIR"
     repeat_region   complement(9537. .9688)
                     /rpt_family="MER5A"
     repeat_region   9797. .9942
                     /rpt_family="MIR"
     repeat_region   complement(10403. .10429)
                     /rpt_family="AT_rich"
     repeat_region   complement(10673. .10705)
                     /rpt_family="(CA)n"
     repeat_region   complement(11460. .11755)
                     /rpt_family="AluSx"
```

```
     repeat_region   11977. .12125
                     /rpt_family="MER5A"
     repeat_region   12556. .13643
                     /rpt_family="L1MD2"
     repeat_region   13618. .13808
                     /rpt_family="L1MD3"
     repeat_region   complement(13997. .14297)
                     /rpt_family="MER4D"
     repeat_region   14327. .14625
                     /rpt_family="AluSg"
     repeat_region   complement(14648. .14915)
                     /rpt_family="MER4D"
     repeat_region   complement(14915. .15147)
                     /rpt_family="MER4D"
     repeat_region   15286. .15404
                     /rpt_family="FLAM_C"
     repeat_region   16169. .16463
                     /rpt_family="AluSx"
     repeat_region   16520. .16570
                     /rpt_family="L2"
     repeat_region   16688. .16904
                     /rpt_family="L2"
     repeat_region   complement(16955. .17147)
                     /rpt_family="MIR"
     repeat_region   complement(17171. .17467)
                     /rpt_family="AluY"
     repeat_region   complement(20101. .20241)
                     /rpt_family="MIR"
     STS             22625. .22788
                     /standard_name="DXS1097"
                     /db_xref="GDB:190032"
     repeat_region   24144. .24435
                     /rpt_family="AluSq"
     repeat_region   complement(25010. .25235)
                     /rpt_family="L2"
     repeat_region   25643. .25771
                     /rpt_family="MER58A"
     repeat_region   complement(26111. .26429)
                     /rpt_family="L2"
     repeat_region   complement(26578. .26636)
                     /rpt_family="(CATA)n"
     repeat_region   complement(26956. .27027)
                     /rpt_family="L1MA9"
     repeat_region   complement(27038. .27188)
                     /rpt_family="L2"
     repeat_region   27203. .27641
                     /rpt_family="L1MA8"
     repeat_region   27642. .28179
                     /rpt_family="L1PA4"
     repeat_region   28194. .28495
                     /rpt_family="AluSx"
     repeat_region   complement(28509. .28660)
                     /rpt_family="L2"
     repeat_region   28665. .28830
                     /rpt_family="AluJb"
     repeat_region   complement(28848. .29503)
                     /rpt_family="L2"

  Query Match             100.0%;  Score 137;  DB 9;  Length 98056;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-63;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91659 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 91718

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91719 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTG 91778

Qy        121 TTTGATGTTAATTAATT 137
              |||||||||||||||||
Db      91779 TTTGATGTTAATTAATT 91795
```

```
RESULT 11
AR220819/c
LOCUS       AR220819               13977 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 60 from patent US 6426186.
ACCESSION   AR220819
VERSION     AR220819.1  GI:23327696
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 13977)
  AUTHORS   Jones,K.A., Volkmuth,W. and Walker,M.G.
  TITLE     Bone remodeling genes
  JOURNAL   Patent: US 6426186-A 60 30-JUL-2002;
FEATURES             Location/Qualifiers
     source          1. .13977
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             85.4%;  Score 117;  DB 6;  Length 13977;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-52;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13306 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 13247

Qy         61 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13246 GCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 13190
```

```
RESULT 12
AC023414
LOCUS       AC023414               212120 bp    DNA     linear   HTG 11-APR-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-767B5 map 14, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC023414
VERSION     AC023414.2  GI:7534009
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-767B5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 11, 2000 this sequence version replaced gi:6970579.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6731
                Center clone name: 767_B_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 185633 bases at least Q40
                Consensus quality: 195115 bases at least Q30
                Consensus quality: 201837 bases at least Q20
                Insert size: 178000; agarose-fp
                Insert size: 209120; sum-of-contigs
                Quality coverage: 5.8 in Q20 bases; agarose-fp
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1043: contig of 1043 bp in length
            *     1044     1143: gap of 100 bp
            *     1144     2224: contig of 1081 bp in length
            *     2225     2324: gap of 100 bp
            *     2325     3345: contig of 1021 bp in length
            *     3346     3445: gap of 100 bp
            *     3446     4502: contig of 1057 bp in length
            *     4503     4602: gap of 100 bp
            *     4603     5756: contig of 1154 bp in length
            *     5757     5856: gap of 100 bp
            *     5857     7097: contig of 1241 bp in length
            *     7098     7197: gap of 100 bp
            *     7198     8236: contig of 1039 bp in length
            *     8237     8336: gap of 100 bp
            *     8337     9353: contig of 1017 bp in length
            *     9354     9453: gap of 100 bp
            *     9454    10705: contig of 1252 bp in length
            *    10706    10805: gap of 100 bp
            *    10806    12332: contig of 1527 bp in length
            *    12333    12432: gap of 100 bp
            *    12433    13877: contig of 1445 bp in length
            *    13878    13977: gap of 100 bp
            *    13978    15201: contig of 1224 bp in length
            *    15202    15301: gap of 100 bp
            *    15302    16968: contig of 1667 bp in length
            *    16969    17068: gap of 100 bp
            *    17069    19060: contig of 1992 bp in length
            *    19061    19160: gap of 100 bp
            *    19161    22639: contig of 3479 bp in length
            *    22640    22739: gap of 100 bp
            *    22740    26376: contig of 3637 bp in length
            *    26377    26476: gap of 100 bp
            *    26477    29275: contig of 2799 bp in length
            *    29276    29375: gap of 100 bp
            *    29376    33265: contig of 3890 bp in length
            *    33266    33365: gap of 100 bp
            *    33366    37012: contig of 3647 bp in length
            *    37013    37112: gap of 100 bp
            *    37113    39390: contig of 2278 bp in length
            *    39391    39490: gap of 100 bp
            *    39491    43690: contig of 4200 bp in length
            *    43691    43790: gap of 100 bp
            *    43791    47621: contig of 3831 bp in length
            *    47622    47721: gap of 100 bp
            *    47722    53852: contig of 6131 bp in length
            *    53853    53952: gap of 100 bp
            *    53953    63422: contig of 9470 bp in length
            *    63423    63522: gap of 100 bp
            *    63523    77361: contig of 13839 bp in length
            *    77362    77461: gap of 100 bp
            *    77462    92465: contig of 15004 bp in length
            *    92466    92565: gap of 100 bp
            *    92566   106951: contig of 14386 bp in length
            *   106952   107051: gap of 100 bp
            *   107052   129678: contig of 22627 bp in length
            *   129679   129778: gap of 100 bp
            *   129779   151427: contig of 21649 bp in length
            *   151428   151527: gap of 100 bp
            *   151528   181838: contig of 30311 bp in length
            *   181839   181938: gap of 100 bp
            *   181939   212120: contig of 30182 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212120
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-767B5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1043
                     /note="assembly_fragment"
     misc_feature    1144. .2224
                     /note="assembly_fragment"
     misc_feature    2325. .3345
                     /note="assembly_fragment"
     misc_feature    3446. .4502
                     /note="assembly_fragment"
     misc_feature    4603. .5756
                     /note="assembly_fragment"
     misc_feature    5857. .7097
                     /note="assembly_fragment"
     misc_feature    7198. .8236
                     /note="assembly_fragment"
     misc_feature    8337. .9353
                     /note="assembly_fragment"
     misc_feature    9454. .10705
                     /note="assembly_fragment"
     misc_feature    10806. .12332
                     /note="assembly_fragment"
     misc_feature    12433. .13877
                     /note="assembly_fragment"
     misc_feature    13978. .15201
                     /note="assembly_fragment"
     misc_feature    15302. .16968
                     /note="assembly_fragment"
     misc_feature    17069. .19060
                     /note="assembly_fragment"
     misc_feature    19161. .22639
                     /note="assembly_fragment"
     misc_feature    22740. .26376
                     /note="assembly_fragment"
     misc_feature    26477. .29275
                     /note="assembly_fragment"
     misc_feature    29376. .33265
                     /note="assembly_fragment"
     misc_feature    33366. .37012
                     /note="assembly_fragment"
     misc_feature    37113. .39390
                     /note="assembly_fragment
                     clone_end:SP6
```

```
                     vector_side:right"
     misc_feature    39491. .43690
                     /note="assembly_fragment"
     misc_feature    43791. .47621
                     /note="assembly_fragment"
     misc_feature    47722. .53852
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    53953. .63422
                     /note="assembly_fragment"
     misc_feature    63523. .77361
                     /note="assembly_fragment"
     misc_feature    77462. .92465
                     /note="assembly_fragment"
     misc_feature    92566. .106951
                     /note="assembly_fragment"
     misc_feature    107052. .129678

  Query Match             47.4%;  Score 65;  DB 2;  Length 212120;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-24;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92401 ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGT 92460

Qy         61 GCACA 65
              |||||
Db      92461 GCACA 92465


RESULT 13
AF070485/c
LOCUS       AF070485               13887 bp    mRNA    linear   MAM 09-DEC-1998
DEFINITION  Canis familiaris dystrophin mRNA, complete cds.
ACCESSION   AF070485
VERSION     AF070485.1  GI:3982750
KEYWORDS    .
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 13887)
  AUTHORS   Carville,K.S., Mann,C.J., Schatzberg,S.J. and Wilton,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) ANRI, Pathology, University of Western
            Australia, Verdun Street, Nedlands, WA 6018, Australia
FEATURES             Location/Qualifiers
     source          1. .13887
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /strain="breed Golden retriever"
                     /db_xref="taxon:9615"
     5'UTR           <1. .289
     CDS             290. .11332
                     /codon_start=1
                     /product="dystrophin"
                     /protein_id="AAC83646.1"
                     /db_xref="GI:3982751"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQKNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIAKYQLGIEKLLDPEDV
                     ATTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPSQVTREEHFQIHHQMHYSQ
                     QITVSLAQGYERAPSFPKPRFKSYAYTQAAYVTTSDPTRSPLPSQHLETPEDKSFGRS
                     LTETEANLDSYQTALEEVLSWLLSAEDALQAQGEISNDVEEVKEQFHTHEGYMMDLTS
                     HQGRVGNVLQLGSQLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHKVL
                     MDLQNQQLKELNDWLTKTEERTRKMEKEPLGPDIEDLKRQVQQHKVLQEDLEQEQVRV
                     NSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRFT
                     EEQCLFSAWLSEKEDAVNKIHTTGFKDQSEVLSNLQKLAVLKTDLEKKKQTMDKLCSL
                     NQDLLSALKNTVVAHKMEAWLDNFAQRWDNLVQKLEKSSAQISQAVTTTQPSLTQTTV
                     METVTMVTTREHILVKHAQEELPPPPPQKKRQIIVDSEIRKRLDVDITELHSWITRSE
                     AVLQSPEFAIYRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSAQALVEQMVNEGVN
                     ADSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIITFYNQLQQLEQMTTTAENWLKT
                     QPTTTSEPTAIKSQLKICKDEINRLSALQPQIERLKIQSIALKEKGQGPMFLDADFVA
                     FTNHFNQVFADVQAREKELQTIFDSLPPMRYQETMSTILTWIQQSETKLSIPQVTVTE
                     YDIMEQRLGELQALQSSLQEQQNGLNYLSTTVKEMSKKAPLSDISRKYQSEFEEIEGR
                     WKKLSSQLVEHCQKLEEQMAKLRKIQNHIKTLKKWITEVDVFLKEEWPALGDSEILKR
                     QLKQCRLLVNDIQTIQPSLNSVNEGAQKMKNEAEPEFAGRLETELRELNTQWDYMCRQ
                     VYARKEALKGGLDKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQTAVEEMKR
                     AKEEAQQKEAKVKLLTESVNSVIAQAPPAAQEALKKELDTLTTNYQWLCTRLNGKCKT
                     LEEVWACWHELLSYLEKANKWLSEVEVKLKTTENISGGAEEIAEVLDSLENLMQHSED
                     NPNQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQEIE
                     KSLHLIQESLSSIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGK
                     ETAQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKS
                     VEQEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKL
                     HYNELGAKVTERKQQLEKCLKLSRKMRKEMNALTEWLAATDMELTKRSAVEGMPSNLD
                     SEVAWGKATQKEIEKQKVHLKSVTEVGEALKTVLGKKEMLVEDKLSLLNSNWIAVTSR
                     AEEWLNLLLEYQKHMETFDQNVDYITNWIIQADALLDESEKKKPQQKEDILKRLKAEM
                     NDIRPKVDSTRDQAANLMANRGDHCRKVVEPKISELNHRFAAISHRIKTGKASIPLKE
                     LEQFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMSEDNEGTVKELLQRGDNLQQRITD
                     ERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEK
                     KLASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRW
                     REIESKFAQFRRLNFAQIHTVHEESVVAMTEDMPLEISYVPSTYLTEITHVSQALSEV
                     EELLNAPDLCAQDFEDLFKQEESLKNIKDSLQQISGRIDIIHNKKTAALHSATPAERA
                     KLQEALSRLDFQWERVNNMYKDRQGRFDRSVEKWRRFHYDMKILNQWLTEAEQFLKKT
                     QIPENWEHAKYKWYLKELQDGIGQRQSVVRVLNATGEEIIQQSSKTDASILQEKLGSL
                     NLRWQEVCKQLAERKKRLEEQKNILSEFQRDVNEFVLWLEEADNVANIPLEPGNEQQL
                     KEKLEQVKLLAEELPLRQGILKQLNETGGTVLVSAPLSPEEQDKLENKLKQTNLQWIK
                     VSRNLPEKQEEIEAHVKDLGQLEEQLNHLLLWLSPIRNQLEIYNQPNQTGPFDIKEIE
                     VAVQAKQPDVEGILSKGQHLYKEKPATQPAKRKLEDLSSDWKVVTQLLQELRAKQPGP
                     APGLTTVRAPPSQTVTLVTQPAVTKETAISKLEMPSSLLLEVPALADFNRAWTELTDW
                     LSLLDRVIKSQRVMVGDLEDINEMIIKQKATLQDLEQRRPQLEELITAAQNLKNKTSN
                     QEARTIITDRIERIQSQWDEVQEHLQNRRLQLTEMLKDSTQWLEAKEEAEQVLGQARA
                     KLESWKEAPYTVDAIQKKITETKQLAKDLRQWQINVDVANDLALKLLRDYSADDTRKV
                     HMITENINASWASIHKRLSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTANVLQD
                     ATHKERLLEDSKGVRELMKQWQDLQGEIEAHTDIYHNLDENGQKVLRSLEGSDDAALL
                     QRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAP
                     IGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPREL
                     PPEERAQNVTRLLRKQAEEVNTQWEKLNVHSADWQRKIDEALERLQELQEATDELDLK
                     LRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEITPLKENVSYVNDLARQLTTLGI
                     QLSPYNLNTLEDLNTRWKLLQVAIEDRIRQLHEAHRDFGPASQHFLSTSVQGPWERAI
                     SPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDL
                     LSLSAACDALDQHNLKQNDQPMDILQVINCLTTIYDRLEQEHNNLVNVPLCVDMCLNW
                     LLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLLH
                     DSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPV
                     LHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYP
                     MVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMETPVTLI
                     NFWPVDSAPASSPQLSHDDTHSRIEHYASRLKKMENSNGSYLNDSISPNESIDDEHLL
                     IQHYWRSLNQESPLSQPRSPAQILISLESEERGELERILADLEGRNRNLQAEYDRLKQ
                     QHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLE
                     SQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSESMGEEDL
                     LSPPQDTSTGLEEVMEQLNHSFPSSRGRNTPGKPMREDTM"
     3'UTR           11333. .>13887
     repeat_region   11789. .11828
                     /rpt_type=tandem
                     /rpt_unit="ca"
ORIGIN

  Query Match             33.6%;  Score 46;  DB 4;  Length 13887;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-14;
  Matches   46;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         72 AGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 117
              |||||||||||||||||||||||||||||||||||||||||||||
Db      13196 AGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCC 13151


RESULT 14
AC108338/c
LOCUS       AC108338               272578 bp    DNA     linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-114P1, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC108338
VERSION     AC108338.4  GI:23101239
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 272578)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 272578)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 272578)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:21737649.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GPMY
                Center clone name: CH230-114P1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 218776 bases at least Q40
                Consensus quality: 222844 bases at least Q30
                Consensus quality: 225163 bases at least Q20
                Estimated insert size: 245370; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   257855: contig of 257855 bp in length
            *   257856   257955: gap of unknown length
            *   257956   261779: contig of 3824 bp in length
            *   261780   261879: gap of unknown length
            *   261880   263114: contig of 1235 bp in length
            *   263115   263214: gap of unknown length
            *   263215   266204: contig of 2990 bp in length
            *   266205   266304: gap of unknown length
            *   266305   272578: contig of 6274 bp in length.
FEATURES             Location/Qualifiers
     source          1. .272578
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-114P1"
     misc_feature    1. .1828
                     /note="wgs_contig"
     misc_feature    257956. .259766
                     /note="wgs_contig"
ORIGIN

  Query Match             33.6%;  Score 46;  DB 2;  Length 272578;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-14;
  Matches   46;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         71 CAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 116
              ||||||||||||||||||||||||||||||||||||||||||||||
Db     239591 CAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 239546


RESULT 15
AC114184
LOCUS       AC114184               279539 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-230F23, *** SEQUENCING IN PROGRESS
            ***, 21 unordered pieces.
ACCESSION   AC114184
VERSION     AC114184.3  GI:23605533
```

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 279539)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 279539)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 279539)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 9, 2002 this sequence version replaced gi:21733949.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GLPV
Center clone name: CH230-230F23
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 212054 bases at least Q40
Consensus quality: 219191 bases at least Q30
Consensus quality: 223518 bases at least Q20
Estimated insert size: 217504; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
---

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	6869: contig of 6869 bp in length
*	6870	6969: gap of unknown length
*	6970	122455: contig of 115486 bp in length
*	122456	122555: gap of unknown length
*	122556	185205: contig of 62650 bp in length
*	185206	185305: gap of unknown length
*	185306	189675: contig of 4370 bp in length
*	189676	189775: gap of unknown length
*	189776	240091: contig of 50316 bp in length
*	240092	240191: gap of unknown length
*	240192	245854: contig of 5663 bp in length
*	245855	245954: gap of unknown length
*	245955	247056: contig of 1102 bp in length
*	247057	247156: gap of unknown length
*	247157	248831: contig of 1675 bp in length
*	248832	248931: gap of unknown length
*	248932	250272: contig of 1341 bp in length
*	250273	250372: gap of unknown length
*	250373	252290: contig of 1918 bp in length
*	252291	252390: gap of unknown length
*	252391	254224: contig of 1834 bp in length
*	254225	254324: gap of unknown length
*	254325	255684: contig of 1360 bp in length
*	255685	255784: gap of unknown length
*	255785	258317: contig of 2533 bp in length
*	258318	258417: gap of unknown length
*	258418	261382: contig of 2965 bp in length
*	261383	261482: gap of unknown length
*	261483	264168: contig of 2686 bp in length
*	264169	264268: gap of unknown length
*	264269	267565: contig of 3297 bp in length
*	267566	267665: gap of unknown length
*	267666	271534: contig of 3869 bp in length
*	271535	271634: gap of unknown length
*	271635	275008: contig of 3374 bp in length
*	275009	275108: gap of unknown length
*	275109	277043: contig of 1935 bp in length

```
            *   277044   277143: gap of unknown length
            *   277144   278161: contig of 1018 bp in length
            *   278162   278261: gap of unknown length
            *   278262   279539: contig of 1278 bp in length.
FEATURES             Location/Qualifiers
     source          1. .279539
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-230F23"
     misc_feature    26239. .33299
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBOG36TJ"
     misc_feature    complement(114677. .115487)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:RWBOG36TV"
     misc_feature    133991. .135460
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    240192. .241530
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             33.6%;  Score 46;  DB 2;  Length 279539;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-14;
  Matches   46;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         71 CAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 116
              ||||||||||||||||||||||||||||||||||||||||||||||
Db     114972 CAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 115017


Search completed: April  6, 2004, 16:30:22
Job time : 581.481 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:22:40 ; Search time 4175.64 Seconds
(without alignments)
10338.454 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	859.8	86.3	98056	9	AC006061	AC006061 Homo sapi
	2	858.2	86.2	212120	2	AC023414	AC023414 Homo sapi
c	3	768.2	77.1	212120	2	AC023414	AC023414 Homo sapi
	4	435	43.7	189131	10	AL645477	AL645477 Mouse DNA
	5	431.4	43.3	272578	2	AC108338	AC108338 Rattus no
	6	412.2	41.4	2463	6	AX817316	AX817316 Sequence
	7	412.2	41.4	4658	9	BC028720	BC028720 Homo sapi
	8	410.8	41.2	5339	6	AX538620	AX538620 Sequence
	9	410.8	41.2	5417	6	AX538619	AX538619 Sequence
	10	410.8	41.2	5462	6	AX538621	AX538621 Sequence
	11	410.8	41.2	8689	6	AX538622	AX538622 Sequence
	12	410.8	41.2	11443	6	AX538624	AX538624 Sequence
	13	410.8	41.2	12057	6	AX538627	AX538627 Sequence
	14	410.8	41.2	12446	9	HSDMDR	X14298 Human mRNA
	15	410.8	41.2	13957	6	AX409637	AX409637 Sequence
	16	410.8	41.2	13957	6	AX538581	AX538581 Sequence
	17	410.8	41.2	13957	9	HUMDYS	M18533 Homo sapien
	18	409.2	41.1	13977	6	AR220819	AR220819 Sequence
c	19	402.4	40.4	279539	2	AC114184	AC114184 Rattus no
	20	396.8	39.8	2148	11	G15848	G15848 human STS C
	21	396.8	39.8	2691	6	AX538618	AX538618 Sequence
	22	374.2	37.6	13815	6	AX306153	AX306153 Sequence
	23	374.2	37.6	13815	6	AX538582	AX538582 Sequence
	24	374.2	37.6	13815	10	MUSDYSA	M68859 Mouse dystr
c	25	374.2	37.6	19307	6	AR093392	AR093392 Sequence
c	26	374.2	37.6	19307	6	AR142592	AR142592 Sequence
	27	368	36.9	13887	4	AF070485	AF070485 Canis fam
	28	350	35.1	350	9	AF213444	AF213444 Homo sapi
c	29	319.8	32.1	56285	2	AC100170	AC100170 Mus muscu
	30	290.4	29.2	2110	9	HUMDMDXX	M92650 Human Duche
	31	282.4	28.4	13575	5	GGDYS	X13369 Chicken mRN
	32	278.8	28.0	2563	9	AK129855	AK129855 Homo sapi
	33	260	26.1	3163	6	E30223	E30223 Shortened d
	34	260	26.1	3172	6	E30222	E30222 Shortened d
	35	260	26.1	4075	6	E30221	E30221 Shortened d
	36	260	26.1	4402	6	E30219	E30219 Shortened d
	37	260	26.1	4402	6	E30220	E30220 Shortened d
	38	252.8	25.4	3747	6	E30218	E30218 Shortened d
	39	218	21.9	218	9	HUMDYST20	M86903 H.sapiens d
	40	189.6	19.0	468	5	AF375546	AF375546 Xenopus l
	41	188.4	18.9	509	5	AF082741	AF082741 Gallus ga
	42	166.2	16.7	1534	5	AF375547	AF375547 Scyliorhi
c	43	151.4	15.2	2148	11	G15848	G15848 human STS C
c	44	151.4	15.2	2563	9	AK129855	AK129855 Homo sapi
c	45	151.4	15.2	2691	6	AX538618	AX538618 Sequence

ALIGNMENTS

RESULT 1
AC006061/c
LOCUS AC006061 98056 bp DNA linear PRI 01-MAY-2002
DEFINITION Homo sapiens X BAC GSHB-19O24 (Genome Systems Human BAC Library) complete sequence.
ACCESSION AC006061
VERSION AC006061.1 GI:4204246
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 98056)
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,

Gorrell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L., Kondejewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 7 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 8 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 30, 1999 this sequence version replaced gi:4176317.
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

FEATURES Location/Qualifiers
source 1. .98056
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="GSHB-19O24"
misc_feature 1. .1999
/note="overlaps bases 151178. .153177 of clone AC117405"
/function="clone overlap"
repeat_region 40. .82
/rpt_family="L2"
repeat_region 378. .507
/rpt_family="L1MD3"
repeat_region complement(518. .570)
/rpt_family="(GGA)n"
repeat_region 571. .729
/rpt_family="L1MC4"
repeat_region 730. .987
/rpt_family="MSTD"
repeat_region complement(991. .1400)
/rpt_family="L2"
repeat_region complement(1649. .1779)
/rpt_family="MIR"
repeat_region complement(1783. .2212)
/rpt_family="LTR16A"
repeat_region complement(3626. .3656)
/rpt_family="(GA)n"
repeat_region complement(4713. .5074)
/rpt_family="THE1B"
repeat_region 5702. .7108
/rpt_family="L2"
repeat_region complement(8670. .8734)
/rpt_family="MLT1G"
repeat_region complement(9069. .9144)
/rpt_family="MLT1G"
repeat_region complement(9254. .9351)
/rpt_family="MIR"
repeat_region complement(9537. .9688)
/rpt_family="MER5A"
repeat_region 9797. .9942
/rpt_family="MIR"
repeat_region complement(10403. .10429)
/rpt_family="AT_rich"
repeat_region complement(10673. .10705)
/rpt_family="(CA)n"
repeat_region complement(11460. .11755)
/rpt_family="AluSx"
repeat_region 11977. .12125
/rpt_family="MER5A"
repeat_region 12556. .13643
/rpt_family="L1MD2"
repeat_region 13618. .13808
/rpt_family="L1MD3"
repeat_region complement(13997. .14297)
/rpt_family="MER4D"
repeat_region 14327. .14625
/rpt_family="AluSg"
repeat_region complement(14648. .14915)
/rpt_family="MER4D"
repeat_region complement(14915. .15147)
/rpt_family="MER4D"
repeat_region 15286. .15404
/rpt_family="FLAM_C"

```
     repeat_region   16169. .16463
                     /rpt_family="AluSx"
     repeat_region   16520. .16570
                     /rpt_family="L2"
     repeat_region   16688. .16904
                     /rpt_family="L2"
     repeat_region   complement(16955. .17147)
                     /rpt_family="MIR"
     repeat_region   complement(17171. .17467)
                     /rpt_family="AluY"
     repeat_region   complement(20101. .20241)
                     /rpt_family="MIR"
     STS             22625. .22788
                     /standard_name="DXS1097"
                     /db_xref="GDB:190032"
     repeat_region   24144. .24435
                     /rpt_family="AluSq"
     repeat_region   complement(25010. .25235)
                     /rpt_family="L2"
     repeat_region   25643. .25771
                     /rpt_family="MER58A"
     repeat_region   complement(26111. .26429)
                     /rpt_family="L2"
     repeat_region   complement(26578. .26636)
                     /rpt_family="(CATA)n"
     repeat_region   complement(26956. .27027)
                     /rpt_family="L1MA9"
     repeat_region   complement(27038. .27188)
                     /rpt_family="L2"
     repeat_region   27203. .27641
                     /rpt_family="L1MA8"
     repeat_region   27642. .28179
                     /rpt_family="L1PA4"
     repeat_region   28194. .28495
                     /rpt_family="AluSx"
     repeat_region   complement(28509. .28660)
                     /rpt_family="L2"
     repeat_region   28665. .28830
                     /rpt_family="AluJb"
     repeat_region   complement(28848. .29503)
                     /rpt_family="L2"

  Query Match             86.3%;  Score 859.8;  DB 9;  Length 98056;
  Best Local Similarity   99.2%;  Pred. No. 1.7e-164;
  Matches  864;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94154 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 94095

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94094 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 94035

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94034 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 93975

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93974 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 93915

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93914 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 93855

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93854 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 93795

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93794 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 93735

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93734 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 93675

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93674 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 93615

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93614 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 93555

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93554 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 93495

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93494 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 93435

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93434 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 93375

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93374 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 93315

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||| |   ||  ||
Db      93314 TAACAGTTATAAAGAAAGATTGTAAACTAAA 93284


RESULT 2
AC023414
LOCUS       AC023414               212120 bp    DNA     linear   HTG 11-APR-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-767B5 map 14, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC023414
VERSION     AC023414.2  GI:7534009
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-767B5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 11, 2000 this sequence version replaced gi:6970579.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L6731
              Center clone name: 767_B_5
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 185633 bases at least Q40
              Consensus quality: 195115 bases at least Q30
              Consensus quality: 201837 bases at least Q20
              Insert size: 178000; agarose-fp
              Insert size: 209120; sum-of-contigs
              Quality coverage: 5.8 in Q20 bases; agarose-fp
              Quality coverage: 4.9 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 31 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1043: contig of 1043 bp in length
          *     1044     1143: gap of 100 bp
          *     1144     2224: contig of 1081 bp in length
          *     2225     2324: gap of 100 bp
          *     2325     3345: contig of 1021 bp in length
          *     3346     3445: gap of 100 bp
          *     3446     4502: contig of 1057 bp in length
          *     4503     4602: gap of 100 bp
          *     4603     5756: contig of 1154 bp in length
          *     5757     5856: gap of 100 bp
          *     5857     7097: contig of 1241 bp in length
          *     7098     7197: gap of 100 bp
          *     7198     8236: contig of 1039 bp in length
          *     8237     8336: gap of 100 bp
          *     8337     9353: contig of 1017 bp in length
          *     9354     9453: gap of 100 bp
          *     9454    10705: contig of 1252 bp in length
          *    10706    10805: gap of 100 bp
          *    10806    12332: contig of 1527 bp in length
          *    12333    12432: gap of 100 bp
          *    12433    13877: contig of 1445 bp in length
          *    13878    13977: gap of 100 bp
          *    13978    15201: contig of 1224 bp in length
          *    15202    15301: gap of 100 bp
          *    15302    16968: contig of 1667 bp in length
          *    16969    17068: gap of 100 bp
          *    17069    19060: contig of 1992 bp in length
          *    19061    19160: gap of 100 bp
          *    19161    22639: contig of 3479 bp in length
          *    22640    22739: gap of 100 bp
          *    22740    26376: contig of 3637 bp in length
          *    26377    26476: gap of 100 bp
          *    26477    29275: contig of 2799 bp in length
          *    29276    29375: gap of 100 bp
          *    29376    33265: contig of 3890 bp in length
          *    33266    33365: gap of 100 bp
          *    33366    37012: contig of 3647 bp in length
          *    37013    37112: gap of 100 bp
          *    37113    39390: contig of 2278 bp in length
          *    39391    39490: gap of 100 bp
          *    39491    43690: contig of 4200 bp in length
          *    43691    43790: gap of 100 bp
          *    43791    47621: contig of 3831 bp in length
          *    47622    47721: gap of 100 bp
          *    47722    53852: contig of 6131 bp in length
          *    53853    53952: gap of 100 bp
          *    53953    63422: contig of 9470 bp in length
          *    63423    63522: gap of 100 bp
          *    63523    77361: contig of 13839 bp in length
          *    77362    77461: gap of 100 bp
          *    77462    92465: contig of 15004 bp in length
          *    92466    92565: gap of 100 bp
          *    92566   106951: contig of 14386 bp in length
          *   106952   107051: gap of 100 bp
          *   107052   129678: contig of 22627 bp in length
          *   129679   129778: gap of 100 bp
          *   129779   151427: contig of 21649 bp in length
          *   151428   151527: gap of 100 bp
          *   151528   181838: contig of 30311 bp in length
          *   181839   181938: gap of 100 bp
          *   181939   212120: contig of 30182 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212120
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-767B5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1043
                     /note="assembly_fragment"
     misc_feature    1144. .2224
                     /note="assembly_fragment"
     misc_feature    2325. .3345
                     /note="assembly_fragment"
     misc_feature    3446. .4502
                     /note="assembly_fragment"
     misc_feature    4603. .5756
                     /note="assembly_fragment"
     misc_feature    5857. .7097
                     /note="assembly_fragment"
     misc_feature    7198. .8236
                     /note="assembly_fragment"
     misc_feature    8337. .9353
                     /note="assembly_fragment"
     misc_feature    9454. .10705
                     /note="assembly_fragment"
     misc_feature    10806. .12332
                     /note="assembly_fragment"
     misc_feature    12433. .13877
                     /note="assembly_fragment"
     misc_feature    13978. .15201
                     /note="assembly_fragment"
     misc_feature    15302. .16968
                     /note="assembly_fragment"
     misc_feature    17069. .19060
                     /note="assembly_fragment"
     misc_feature    19161. .22639
                     /note="assembly_fragment"
     misc_feature    22740. .26376
                     /note="assembly_fragment"
     misc_feature    26477. .29275
                     /note="assembly_fragment"
     misc_feature    29376. .33265
                     /note="assembly_fragment"
     misc_feature    33366. .37012
                     /note="assembly_fragment"
     misc_feature    37113. .39390
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
```

```
     misc_feature    39491. .43690
                     /note="assembly_fragment"
     misc_feature    43791. .47621
                     /note="assembly_fragment"
     misc_feature    47722. .53852
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    53953. .63422
                     /note="assembly_fragment"
     misc_feature    63523. .77361
                     /note="assembly_fragment"
     misc_feature    77462. .92465
                     /note="assembly_fragment"
     misc_feature    92566. .106951
                     /note="assembly_fragment"
     misc_feature    107052. .129678

  Query Match             86.2%;  Score 858.2;  DB 2;  Length 212120;
  Best Local Similarity   99.1%;  Pred. No. 3.1e-164;
  Matches  863;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180416 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 180475

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180476 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 180535

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180536 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180595

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     180596 GTAGCTTGACATTTTTCCATGTTAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 180655

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180656 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 180715

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180716 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 180775

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180776 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 180835

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180836 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 180895

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180896 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 180955

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180956 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 181015

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181016 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 181075

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181076 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 181135

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181136 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 181195

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181196 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 181255

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||| |   ||   ||
Db     181256 TAACAGTTATAAAGAAAGATTGTAAACTAAA 181286


RESULT 3
AC023414/c
LOCUS       AC023414               212120 bp    DNA     linear   HTG 11-APR-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-767B5 map 14, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC023414
VERSION     AC023414.2  GI:7534009
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-767B5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 11, 2000 this sequence version replaced gi:6970579.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6731
                Center clone name: 767_B_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 185633 bases at least Q40
                Consensus quality: 195115 bases at least Q30
                Consensus quality: 201837 bases at least Q20
                Insert size: 178000; agarose-fp
                Insert size: 209120; sum-of-contigs
                Quality coverage: 5.8 in Q20 bases; agarose-fp
```

```
    Quality coverage: 4.9 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available.and the accession number will
* be preserved.
*        1     1043: contig of 1043 bp in length
*     1044     1143: gap of 100 bp
*     1144     2224: contig of 1081 bp in length
*     2225     2324: gap of 100 bp
*     2325     3345: contig of 1021 bp in length
*     3346     3445: gap of 100 bp
*     3446     4502: contig of 1057 bp in length
*     4503     4602: gap of 100 bp
*     4603     5756: contig of 1154 bp in length
*     5757     5856: gap of 100 bp
*     5857     7097: contig of 1241 bp in length
*     7098     7197: gap of 100 bp
*     7198     8236: contig of 1039 bp in length
*     8237     8336: gap of 100 bp
*     8337     9353: contig of 1017 bp in length
*     9354     9453: gap of 100 bp
*     9454    10705: contig of 1252 bp in length
*    10706    10805: gap of 100 bp
*    10806    12332: contig of 1527 bp in length
*    12333    12432: gap of 100 bp
*    12433    13877: contig of 1445 bp in length
*    13878    13977: gap of 100 bp
*    13978    15201: contig of 1224 bp in length
*    15202    15301: gap of 100 bp
*    15302    16968: contig of 1667 bp in length
*    16969    17068: gap of 100 bp
*    17069    19060: contig of 1992 bp in length
*    19061    19160: gap of 100 bp
*    19161    22639: contig of 3479 bp in length
*    22640    22739: gap of 100 bp
*    22740    26376: contig of 3637 bp in length
*    26377    26476: gap of 100 bp
*    26477    29275: contig of 2799 bp in length
*    29276    29375: gap of 100 bp
*    29376    33265: contig of 3890 bp in length
*    33266    33365: gap of 100 bp
*    33366    37012: contig of 3647 bp in length
*    37013    37112: gap of 100 bp
*    37113    39390: contig of 2278 bp in length
*    39391    39490: gap of 100 bp
*    39491    43690: contig of 4200 bp in length
*    43691    43790: gap of 100 bp
*    43791    47621: contig of 3831 bp in length
*    47622    47721: gap of 100 bp
*    47722    53852: contig of 6131 bp in length
*    53853    53952: gap of 100 bp
*    53953    63422: contig of 9470 bp in length
*    63423    63522: gap of 100 bp
*    63523    77361: contig of 13839 bp in length
*    77362    77461: gap of 100 bp
*    77462    92465: contig of 15004 bp in length
*    92466    92565: gap of 100 bp
*    92566   106951: contig of 14386 bp in length
*   106952   107051: gap of 100 bp
*   107052   129678: contig of 22627 bp in length
*   129679   129778: gap of 100 bp
*   129779   151427: contig of 21649 bp in length
*   151428   151527: gap of 100 bp
*   151528   181838: contig of 30311 bp in length
*   181839   181938: gap of 100 bp
*   181939   212120: contig of 30182 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212120
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-767B5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1043
                     /note="assembly_fragment"
     misc_feature    1144. .2224
                     /note="assembly_fragment"
     misc_feature    2325. .3345
                     /note="assembly_fragment"
     misc_feature    3446. .4502
                     /note="assembly_fragment"
     misc_feature    4603. .5756
                     /note="assembly_fragment"
     misc_feature    5857. .7097
                     /note="assembly_fragment"
     misc_feature    7198. .8236
                     /note="assembly_fragment"
     misc_feature    8337. .9353
                     /note="assembly_fragment"
     misc_feature    9454. .10705
                     /note="assembly_fragment"
     misc_feature    10806. .12332
                     /note="assembly_fragment"
     misc_feature    12433. .13877
                     /note="assembly_fragment"
     misc_feature    13978. .15201
                     /note="assembly_fragment"
     misc_feature    15302. .16968
                     /note="assembly_fragment"
     misc_feature    17069. .19060
                     /note="assembly_fragment"
     misc_feature    19161. .22639
                     /note="assembly_fragment"
     misc_feature    22740. .26376
                     /note="assembly_fragment"
     misc_feature    26477. .29275
                     /note="assembly_fragment"
     misc_feature    29376. .33265
                     /note="assembly_fragment"
     misc_feature    33366. .37012
                     /note="assembly_fragment"
     misc_feature    37113. .39390
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    39491. .43690
                     /note="assembly_fragment"
     misc_feature    43791. .47621
                     /note="assembly_fragment"
     misc_feature    47722. .53852
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    53953. .63422
                     /note="assembly_fragment"
     misc_feature    63523. .77361
                     /note="assembly_fragment"
     misc_feature    77462. .92465
                     /note="assembly_fragment"
     misc_feature    92566. .106951
                     /note="assembly_fragment"
     misc_feature    107052. .129678

Query Match             77.1%;  Score 768.2;  DB 2;  Length 212120;
Best Local Similarity   94.6%;  Pred. No. 5.4e-146;
Matches  828;  Conservative    0;  Mismatches   43;  Indels    4;  Gaps    3;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              |||||||||||||||||||||||||||||| |||||||||||||||||| ||||||| ||
```

```
Db      13463 GTGGTTTGATTGATAGTAAAAAAAATGTTTGTTAATACAAGTAGAGAGTGAGTAATCTAT 13404

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120

Db      13403 CTATCTCTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCGTGTTGTGAT 13344

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGG--AACCTGA 178

Db      13343 AGAAATCTTGTGCTTATATAGGGAATTCTTTTGATATATATTCACATTGGGCAACCTGAG 13284

Qy        179 ATGTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATAC 238

Db      13283 ATGTAGCTTGACATTTTTCCATGTTCACCCCAGTAGCGTGATCCCACATTAAGCTGATAC 13224

Qy        239 TAACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGA 298

Db      13223 TCACAAACAACGTGTAATGGTTTCATTAATAAGGCTTTGTTTTTTCGTGGAAAGTGGTGA 13164

Qy        299 AAAATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAA 358

Db      13163 AAAATCACACCTTGTTGTGTACCCCCTAGATGCAGTTTGTGTGTTGTTTTCACCCAGAAA 13104

Qy        359 TGGGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTAT-CTGCACCTTTT 417

Db      13103 TGGGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATGCTGCACCTTTT 13044

Qy        418 GTAAAGT-CTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCA 476

Db      13043 GTAAAGTGCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCA 12984

Qy        477 CATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGA 536

Db      12983 CATGGCAGATGATTTGGGCAGAGCGATGGAGTCTTTAGTATCAGTCATGACAGATGAAGA 12924

Qy        537 AGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATAC 596

Db      12923 AGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCACAC 12864

Qy        597 AACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGT 656

Db      12863 AACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATATATATTTTTGTGAAGGGT 12804

Qy        657 AGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATG 716

Db      12803 AGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATG 12744

Qy        717 GCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAAT 776

Db      12743 GCAGGTTTTACACGTATATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAAT 12684

Qy        777 CTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAA 836

Db      12683 CTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAA 12624

Qy        837 TTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871

Db      12623 TTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 12589
```

```
RESULT 4
AL645477
LOCUS       AL645477              189131 bp    DNA     linear   ROD 29-JUN-2002
DEFINITION  Mouse DNA sequence from clone RP23-155L18 on chromosome X, complete
            sequence.
ACCESSION   AL645477
VERSION     AL645477.8  GI:20338482
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189131)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 29, 2002 this sequence version replaced gi:17384537.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-155L18 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------.
FEATURES             Location/Qualifiers
     source          1. .189131
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-155L18"
                     /clone_lib="RPCI-23"
ORIGIN
```

Query Match 43.7%; Score 435; DB 10; Length 189131;
Best Local Similarity 84.7%; Pred. No. 1.8e-78;
Matches 577; Conservative 0; Mismatches 80; Indels 24; Gaps 7;

```
Qy        192 TTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTAACAAACAACGT 251

Db     113722 TTCTCCCCATCAAATGACACCATGCTGATCCAGTATTAAGCTAATACTAAC-----ACCA 113776

Qy        252 GTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAAAATCAAACCTT 311

Db     113777 TGCAATGCTTCATTAACAAGGATTTGCTTCTTGCTAGAAATGGGTAAAAA--CGGACTGT 113834

Qy        312 GTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATGGGGAATGATTT 371

Db     113835 GGTCTGTATACCTTCAATGCAGCTTATGTGTTGTCTTTTCC--TGAAATGGTAATGACTC 113892

Qy        372 CCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTAAAGTCTGTCTT 431

Db     113893 CCAATAGTGGCAACCAGGGGTACAATACT-------TGCACACTTTGTAAA----CTCTT 113941

Qy        432 TCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTT 491

Db     113942 TCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTT 114001

Qy        492 GGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAAT 551

Db     114002 GGGCAGAGCGATGGAGTCCTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAAT 114061

Qy        552 GTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAG 611
```

```
Db     114062 GTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAG 114121

Qy        612 ACAGTAAGAGTTTACAAGAAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACT 670
              |||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||
Db     114122 ACAGTAAGAGTTTACAAGAAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACT 114181

Qy        671 GTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114182 GTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACG 114241

Qy        731 TCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAAT 790
              |||||||||||||||||||||| ||||||   ||||||||||||||||||||||||||
Db     114242 TCTATGCAATTGTACAAAAAAGTTAAAAGAAA---ACATGTAAAATCTTGATAGCTAAAT 114298

Qy        791 AACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTAT 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114299 AACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTAT 114358

Qy        851 AAAGAAAGAATTATAAAGGAA 871
              |||||||| |   ||   ||
Db     114359 AAAGAAAGATTGTAAACTAAA 114379
```

RESULT 5
AC108338

LOCUS	AC108338 272578 bp DNA linear HTG 08-OCT-2002
DEFINITION	Rattus norvegicus clone CH230-114P1, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION	AC108338
VERSION	AC108338.4 GI:23101239
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 272578)
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 272578)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 272578)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Sep 18, 2002 this sequence version replaced gi:21737649. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table. -------------- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu -------------- Project Information Center project name: GPMY Center clone name: CH230-114P1 -------------- Summary Statistics Assembly program: Phrap; version 0.990329 Consensus quality: 218776 bases at least Q40 Consensus quality: 222844 bases at least Q30 Consensus quality: 225163 bases at least Q20 Estimated insert size: 245370; sum-of-contigs estimation Quality coverage: 4x in Q20 bases; sum-of-contigs estimation -------------- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 5 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 257855: contig of 257855 bp in length * 257856 257955: gap of unknown length * 257956 261779: contig of 3824 bp in length * 261780 261879: gap of unknown length * 261880 263114: contig of 1235 bp in length * 263115 263214: gap of unknown length * 263215 266204: contig of 2990 bp in length

```
           *   266205   266304: gap of unknown length
           *   266305   272578: contig of 6274 bp in length.
FEATURES             Location/Qualifiers
     source          1. .272578
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-114P1"
     misc_feature    1. .1828
                     /note="wgs_contig"
     misc_feature    257956. .259766
                     /note="wgs_contig"
ORIGIN
```

Query Match 43.3%; Score 431.4; DB 2; Length 272578;
Best Local Similarity 80.5%; Pred. No. 9e-78;
Matches 611; Conservative 0; Mismatches 121; Indels 27; Gaps 8;

```
Qy        117 TGATAGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAA-CC 175
              ||||| | ||| | |  ||||| | | ||||| || | ||   |  |||| |
Db     237426 TGATAAGAGCCTTATACCCGTCTATGAAGTCCCTTTGAGATGTGTTAGGAAGAGGAATGC 237485

Qy        176 TGAATGTAGC---TTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGC 232
              |||||| | |   ||    |||| ||    |||  |   |  |||||||| ||||||||
Db     237486 TGAATGCATCAATTTACATTTTTCCCCATCAAATGACGCCACGCTGATCCAGCATTAAGC 237545

Qy        233 TGATACTAACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAAC 292
              | |||| |||       |     | |||||| ||| ||||||||||| ||  || ||||
Db     237546 TAATACCAAC-----TCCATGCAATGCTTCACTAACAAGGCTTTGCTGCTCGCTAGAAAT 237600

Qy        293 TGGTGAAAAATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACC 352
              |   ||||    ||  |||| || | ||| || |||||||||||||||||||   ||
Db     237601 GGT---AAAACTGGACTGTGTTCTGCATACCTTCAATGCAGCTTCTGTGTTGTCTTTTCC 237657

Qy        353 CAGAAATGGGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCAC 412
                | |   | |||||| ||    |||||  |       |       |  ||||||
Db     237658 ---AGAAATGTAATGATTCCCATTGTGGCAACCA--------GGGCTACAATGCCTGCAC 237706

Qy        413 CTTTTGTAAAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTT 472
              |||||||  ||| | ||| | |||||||||||| ||||||||||||||||||  ||||
Db     237707 ACTTTGTAACCTCTTTTTTTTTCTCTCTTTGTTTT-CAGGACACAATGTAGGAAGCCTTT 237765

Qy        473 TCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATG 532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237766 TCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTTTCAGTCATGATAGATG 237825

Qy        533 AAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTC 592
              |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db     237826 AAGAAGGAGCAGAATAAATGTTTTACAACTCCTGACCCCCGCATGGTTTTTATAATATTC 237885

Qy        593 ATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAA 652
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     237886 GTACAACAAAGAGGATTAGACAGTAAGCGTTTACAAGAAATAAATCTATATTTTTGTGAA 237945

Qy        653 GGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAAC 712
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     237946 GGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAAC 238005

Qy        713 AATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTA 772
              |||||||||||||||||||||||||||||||||||||||||| ||||||   |||||||
Db     238006 AATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTAAAAGAAA---ACATGTA 238062

Qy        773 AAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTA 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238063 AAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTA 238122

Qy        833 AAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||||||||||| |   ||  ||
Db     238123 AAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 238161
```

RESULT 6
AX817316

```
LOCUS       AX817316               2463 bp    DNA     linear   PAT 10-DEC-2003
DEFINITION  Sequence 64 from Patent WO02081517.
ACCESSION   AX817316
VERSION     AX817316.1  GI:39722703
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Decristofaro,M.F., Padigaru,M., Miller,C., Tchernev,V., Zhong,H.,
            Zhong,M., Anderson,D., Ballinger,R., Gerlach,V., Spytek,K.A.,
            Rastelli,L., Kekuda,R., Guo,X., Zerhusen,B., Andrew,D., Mezes,P.,
            Patturajan,M., Burgess,C.E., Eisen,A., Wolenc,A., Baumgartner,J.,
            Shimkets,R.A., Gusev,V., Vernet,C.A., Taupier,R.J., Pena,C.,
            Shenoy,S., Li,L., Casman,S., Bolgog,F., Fernandes,E., Smithson,G.,
            Malyankar,U., Taillon,B. and Liu,X.
  TITLE     Novel polypeptides and nucleic acids encoded thereby
  JOURNAL   Patent: WO 02081517-A 64 17-OCT-2002;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2463
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             <3. .2108
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE84363.1"
                     /db_xref="GI:39722704"
                     /translation="WASRHTPAPSGGSAPSPARPRRLLRSAFSCELGRRRRRRSTFGE
                     PGGSGKLTPPLVPTLDRGALAAMREQLKGHETQTTCWDHPKMTELYQSLADLNNVRFS
                     AYRTAMKLRRLQKALCWDLLSLSAACDALDQHNLKQNDQPMDILQIIINCLTTIYDRLE
                     QEHNNLVNVPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRNLFK
                     QVASSTGFCDQRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEA
                     ALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDI
                     CQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGY
                     LPVQTVLEGDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNG
                     SYLNDSISPNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERIL
                     ADLEEENRNLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQH
                     KGRLEARMQILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQP
                     MLLRVVGSQTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADD
                     LGRAMESLVSVMTDEEGAE"
ORIGIN
```

Query Match 41.4%; Score 412.2; DB 6; Length 2463;
Best Local Similarity 97.0%; Pred. No. 1.5e-73;
Matches 420; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy        439 CTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 498
              | | |||    ||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CCTAGTTCAAGAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 2057

Qy        499 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 2117

Qy        559 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 2177

Qy        619 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 2237

Qy        679 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2238 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 2297

Qy        739 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCC 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2298 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCC 2357
```

```
Qy        799 ATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2358 ATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAG 2417

Qy        859 AATTATAAAGGAA 871
              | |  ||  ||
Db       2418 ATTGTAAACTAAA 2430
```

RESULT 7
BC028720

```
LOCUS       BC028720                4658 bp    mRNA    linear   PRI 25-AUG-2003
DEFINITION  Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker
            types), transcript variant Dp71b, mRNA (cDNA clone IMAGE:4822807),
            complete cds.
ACCESSION   BC028720
VERSION     BC028720.1  GI:20379675
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4658)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4658)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 46 Row: j Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 5032296
            This clone has the following problem: The cds is short compared to
            the longest cds in the locus.
```

```
FEATURES             Location/Qualifiers
     source          1. .4658
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4822807"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     gene            1. .4658
                     /gene="DMD"
                     /note="synonyms: BMD, DXS142, DXS164, DXS206, DXS230,
                     DXS239, DXS268, DXS269, DXS270, DXS272"
                     /db_xref="LocusID:1756"
                     /db_xref="MIM:300377"
     CDS             130. .2037
                     /codon_start=1
                     /product="dystrophin, Dp71b isoform"
                     /protein_id="AAH28720.1"
                     /db_xref="GI:20379676"
                     /db_xref="LocusID:1756"
                     /translation="MREQLKGHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKL
                     RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVN
                     VPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGF
                     CDQRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMR
                     LEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSG
                     RVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLE
                     GDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSIS
                     PNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENR
                     NLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARM
                     QILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGS
                     QTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADDLGRAMESL
                     VSVMTDEEGAE"
     misc_feature    844. .981
                     /note="ZZ; Region: Zinc finger, ZZ type. Zinc finger
                     present in dystrophin, CBP/p300. ZZ in dystrophin binds
                     calmodulin Putative zinc finger"
                     /db_xref="CDD:pfam00569"
ORIGIN
```

```
  Query Match             41.4%;  Score 412.2;  DB 9;  Length 4658;
  Best Local Similarity   97.0%;  Pred. No. 1.3e-73;
  Matches  420;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy        439 CTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 498
              | | |||     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1927 CCTAGTTCAAGAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 1986

Qy        499 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1987 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 2046

Qy        559 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2047 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 2106

Qy        619 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2107 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 2166

Qy        679 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2167 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 2226

Qy        739 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCC 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2227 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCC 2286

Qy        799 ATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2287 ATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAG 2346
```

```
Qy        859 AATTATAAAGGAA 871
              | |   ||   ||
Db       2347 ATTGTAAACTAAA 2359


RESULT 8
AX538620
LOCUS       AX538620               5339 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 40 from Patent WO0229056.
ACCESSION   AX538620
VERSION     AX538620.1  GI:25271166
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 40 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .5339
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410.8;  DB 6;  Length 5339;
  Best Local Similarity   98.3%;  Pred. No. 2.5e-73;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4551 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4610

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4611 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4730

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4731 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4790

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4791 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4850

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4851 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 4910

Qy        870 AA 871
              ||
Db       4911 AA 4912


RESULT 9
AX538619
LOCUS       AX538619               5417 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 39 from Patent WO0229056.
ACCESSION   AX538619
VERSION     AX538619.1  GI:25271163
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 39 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .5417
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410.8;  DB 6;  Length 5417;
  Best Local Similarity   98.3%;  Pred. No. 2.5e-73;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4628

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4629 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4688

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4689 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4748

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4749 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4808

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4868

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4869 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4928

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 4988

Qy        870 AA 871
              ||
Db       4989 AA 4990


RESULT 10
AX538621
LOCUS       AX538621               5462 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 41 from Patent WO0229056.
ACCESSION   AX538621
VERSION     AX538621.1  GI:25271168
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 41 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .5462
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
```

```
ORIGIN

  Query Match           41.2%;  Score 410.8;  DB 6;  Length 5462;
  Best Local Similarity  98.3%;  Pred. No. 2.5e-73;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4673

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4733

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4734 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4793

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4794 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4853

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4854 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4913

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4914 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4973

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4974 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 5033

Qy        870 AA 871
              ||
Db       5034 AA 5035


RESULT 11
AX538622
LOCUS       AX538622                8689 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 42 from Patent WO0229056.
ACCESSION   AX538622
VERSION     AX538622.1  GI:25271171
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 42 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .8689
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match           41.2%;  Score 410.8;  DB 6;  Length 8689;
  Best Local Similarity  98.3%;  Pred. No. 2.3e-73;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 6044

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 6104

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6105 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 6164

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6165 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 6224

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6225 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 6284

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6285 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 6344

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6345 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 6404

Qy        870 AA 871
              ||
Db       6405 AA 6406


RESULT 12
AX538624
LOCUS       AX538624               11443 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 44 from Patent WO0229056.
ACCESSION   AX538624
VERSION     AX538624.1  GI:25271175
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 44 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .11443
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match           41.2%;  Score 410.8;  DB 6;  Length 11443;
  Best Local Similarity  98.3%;  Pred. No. 2.2e-73;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8739 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 8798

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 8858

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8859 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 8918

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8919 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 8978

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8979 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 9038
```

```
Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9039 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 9098

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db       9099 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 9158

Qy        870 AA 871
              ||
Db       9159 AA 9160
```

```
RESULT 13
AX538627
LOCUS       AX538627               12057 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 47 from Patent WO0229056.
ACCESSION   AX538627
VERSION     AX538627.1  GI:25271181
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 47 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .12057
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN
```

```
  Query Match           41.2%;  Score 410.8;  DB 6;  Length 12057;
  Best Local Similarity  98.3%;  Pred. No. 2.2e-73;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11214 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11273

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11274 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11333

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11334 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11393

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11394 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11453

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11454 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11513

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11514 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11573

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db      11574 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11633

Qy        870 AA 871
              ||
Db      11634 AA 11635
```

```
RESULT 14
HSDMDR
LOCUS       HSDMDR                 12446 bp    mRNA    linear   PRI 12-SEP-1993
DEFINITION  Human mRNA for dystrophin.
ACCESSION   X14298
VERSION     X14298.1  GI:30845
KEYWORDS    Dmd gene; Duchenne muscular dystrophy; dystrophin.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 12446)
  AUTHORS   Rosenthal,A., Speer,A., Billwitz,H., Cross,G.S., Forrest,S.M. and
            Davies,K.E.
  TITLE     Two human cDNA molecules coding for the Duchenne muscular dystrophy
            (DMD) locus are highly homologous
  JOURNAL   Nucleic Acids Res. 17 (13), 5391 (1989)
  MEDLINE   89345106
   PUBMED   2668885
REFERENCE   2  (bases 1 to 12446)
  AUTHORS   Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1989) Rosenthal A., Akademie der Wissenschaften
            der DDr, Zentralinstitut fuer Molekularbiologie, Robert-Roessle
            Str.10, 1115 Berlin Buch, DDR
COMMENT     see also M18533 and M20250 for Dmd seqs.; discrepancies compared to
            M18533 cDNA were located at x14298 pos. 496, 1772, 1965, 2449,
            3687, 4229, 4504, 5075, 5332, 5630 and 7194.
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X chromosomal, Xp21."
                     /tissue_type="muscle"
                     /dev_stage="fetal and adult."
     CDS             99. .11156
                     /note="unnamed protein product; dystrophin (AA 1 - 3685)"
                     /codon_start=1
                     /protein_id="CAA32479.1"
                     /db_xref="GI:30846"
                     /db_xref="GOA:P11532"
                     /db_xref="SWISS-PROT:P11532"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQPTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDV
                     DTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQ
                     QITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL
                     MESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTAH
                     QGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLM
                     DLQNQKLKELNDWLTKTEERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVN
                     SLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTE
                     EQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSIK
                     QDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVM
                     ETVTTVTTREQILVKHAQEELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEA
                     VLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSGQALVEQMVNEGVNA
                     DSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQ
                     PTTPSEPTAIKSQLKICKDEVNRLSGLQPQIERLKIQSIALKEKGQGPMFLDADFVAF
                     TNHFKQVFSDVQAREKELQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDY
                     EIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWK
                     KLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL
                     KQCRLLVSDIQTIQPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVY
                     ARKEALKGGLEKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQKAFEEMKRAK
                     EEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNGKCKTLE
                     EVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNP
                     NQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETENS
                     LHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEA
                     AQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKSVE
                     QEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHY
                     NELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSE
                     VAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIAVTSRAE
                     EWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELND
```

IRPKVDSTRDQAANLMANHGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELE
QFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQRGDNLQQRITDER
KSEEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKL
ASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWRE
IESKFAQFRRLNFAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQ
LLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSGRIDIIHSKKTAALQSATPVERVKL
QEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI
PENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNL
RWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKE
KLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKLKQTNLQWIKVS
RALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFD
VKETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRA
KQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWT
ELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLEELITAAQNLK
NKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVL
GQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSAD
DTRKVHMITENINASWRSIHKRVSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTA
NVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQKILRSLEGSD
DAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDEL
SRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQ
EPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNLHSADWQRKIDETLERLQELQEATD
ELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQL
TTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP
WERAISPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKA
LCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVD
MCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRL
GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSM
VWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGH
KMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
PVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESID
DEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEY
DRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDH
NKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSDSM
GEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM"

ORIGIN

Query Match 41.2%; Score 410.8; DB 9; Length 12446;
Best Local Similarity 98.3%; Pred. No. 2.2e-73;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11143 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11202

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11203 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11262

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11263 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11322

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11323 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11382

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11383 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11442

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11443 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11502

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11503 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11562

Qy        870 AA 871
              ||
Db      11563 AA 11564
```

RESULT 15
AX409637
LOCUS AX409637 13957 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2284 from Patent WO0229103.
ACCESSION AX409637
VERSION AX409637.1 GI:21442342
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2284 11-APR-2002;
GENE LOGIC INC (US)
FEATURES Location/Qualifiers
source 1. .13957
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. M18533"

ORIGIN

Query Match 41.2%; Score 410.8; DB 6; Length 13957;
Best Local Similarity 98.3%; Pred. No. 2.2e-73;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

Search completed: April 6, 2004, 13:02:08
Job time : 4177.64 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 02:56:20 ; Search time 471.188 Seconds
(without alignments)
8979.866 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	996	100.0	996	6	ABK86497	Abk86497 Human Apo
	2	996	100.0	1230	6	ABK86462	Abk86462 Human Apo
	3	996	100.0	1234	6	ABK86463	Abk86463 Human Apo
	4	435	43.7	1044	6	ABK86464	Abk86464 Mouse gen
	5	412.2	41.4	2463	7	ABT33375	Abt33375 NOVX DNA
	6	410.8	41.2	5339	6	ABK81998	Abk81998 DNA encod
	7	410.8	41.2	5417	6	ABK81997	Abk81997 DNA encod
	8	410.8	41.2	5462	6	ABK81999	Abk81999 DNA encod
	9	410.8	41.2	8689	6	ABK82000	Abk82000 DNA encod
	10	410.8	41.2	11443	6	ABK82002	Abk82002 DNA encod
	11	410.8	41.2	12923	1	AAN90338	Aan90338 Sequence
	12	410.8	41.2	13957	6	ABK81959	Abk81959 cDNA enco
	13	410.8	41.2	13957	6	ABT10904	Abt10904 Human bre
	14	410.8	41.2	13957	6	ABN95786	Abn95786 Gene #228
	15	410.8	41.2	13957	6	ABS69900	Abs69900 Human dys
	16	409.2	41.1	13977	6	ABS70403	Abs70403 Human bon
	17	396.8	39.8	2691	6	ABK81996	Abk81996 Human dys
	18	374.2	37.6	13815	2	AAV18885	Aav18885 Mus muscu
	19	374.2	37.6	13815	6	ABK81960	Abk81960 cDNA enco
	20	374.2	37.6	13815	6	ABI99799	Mouse isc
c	21	374.2	37.6	19307	2	AAT27558	Aat27558 Shuttle v
	22	260	26.1	3163	3	AAZ48571	Aaz48571 A rod sho
	23	260	26.1	3172	3	AAZ48570	Aaz48570 A rod sho
	24	260	26.1	4075	3	AAZ48569	Aaz48569 A rod sho
	25	260	26.1	4402	3	AAZ48568	Aaz48568 A rod sho
	26	260	26.1	4402	3	AAZ48567	Aaz48567 A rod sho
	27	252.8	25.4	3747	3	AAZ48566	Aaz48566 A rod sho
c	28	158	15.9	158	6	ABK86469	Abk86469 Human apo
c	29	151.4	15.2	455	9	ADD32504	Add32504 Human mit
c	30	151.4	15.2	2691	6	ABK81996	Abk81996 Human dys
c	31	151.4	15.2	8689	6	ABK82000	Abk82000 DNA encod
c	32	151.4	15.2	11443	6	ABK82002	Abk82002 DNA encod
c	33	151.4	15.2	13957	6	ABK81959	Abk81959 cDNA enco
c	34	151.4	15.2	13957	6	ABT10904	Abt10904 Human bre
c	35	151.4	15.2	13957	6	ABN95786	Abn95786 Gene #228
c	36	151.4	15.2	13957	6	ABS69900	Abs69900 Human dys
c	37	146.4	14.7	200	6	ABK86468	Abk86468 Human dys
	38	146.4	14.7	2005	7	ABT33376	Abt33376 NOVX DNA
c	39	140.4	14.1	13977	6	ABS70403	Abs70403 Human bon
	40	137	13.8	137	6	ABK86496	Abk86496 Human Apo
c	41	108.4	10.9	959	6	ABN74601	Abn74601 Bovine em
c	42	101.4	10.2	13815	2	AAV18885	Aav18885 Mus muscu
c	43	101.4	10.2	13815	6	ABK81960	Abk81960 cDNA enco
c	44	101.4	10.2	13815	6	ABI99799	Abi99799 Mouse isc
	45	101.4	10.2	19307	2	AAT27558	Aat27558 Shuttle v

ALIGNMENTS

RESULT 1
ABK86497
ID ABK86497 standard; cDNA; 996 BP.
XX
AC ABK86497;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human Apo-dystrophin-4 cDNA.
XX
KW Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .996
FT /*tag= a
FT /product= "Apo-dystrophin-4 protein appearing as
FT AAU98739"
FT /partial
FT /note= "No start or stop codon shown"
FT /transl_except= (pos:7. .9,aa:Xaa)
FT /transl_except= (pos:49. .51,aa:Xaa)
FT /transl_except= (pos:187. .189,aa:Xaa)
FT /transl_except= (pos:202. .204,aa:Xaa)
FT /transl_except= (pos:217. .219,aa:Xaa)
FT /transl_except= (pos:253. .255,aa:Xaa)
FT /transl_except= (pos:265. .270,aa:Xaa-Xaa)
FT /transl_except= (pos:394. .396,aa:Xaa)
FT /transl_except= (pos:547. .549,aa:Xaa)
FT /transl_except= (pos:565. .567,aa:Xaa)
FT /transl_except= (pos:616. .618,aa:Xaa)
FT /transl_except= (pos:649. .651,aa:Xaa)
FT /transl_except= (pos:682. .684,aa:Xaa)
FT /transl_except= (pos:709. .711,aa:Xaa)
FT /transl_except= (pos:790. .792,aa:Xaa)
FT /transl_except= (pos:841. .843,aa:Xaa)
FT /transl_except= (pos:850. .852,aa:Xaa)
FT /transl_except= (pos:880. .882,aa:Xaa)
FT /transl_except= (pos:907. .909,aa:Xaa)
FT /transl_except= (pos:952. .954,aa:Xaa)
FT /transl_except= (pos:982. .984,aa:Xaa)
FT /transl_except= (pos:988. .990,aa:Xaa)
FT /note= "Xaa= unknown, encoded by in frame stop codon"
XX

PN GB2368064-A.
XX
PD 24-APR-2002.
XX
PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
DR P-PSDB; AAU98739.
XX
PT New human regulatory polynucleotide, useful for treating disorders
PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Claim 5; Page 170-172; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is the cDNA sequence for human apo-dystrophin-4
CC containing a plurality of stop codons some of which may be read through
CC due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;

Query Match 100.0%; Score 996; DB 6; Length 996;
Best Local Similarity 100.0%; Pred. No. 2.4e-211;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTGATATATATTTACATTGGGAACCTGAAT 180

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
```

RESULT 2
ABK86462
ID ABK86462 standard; cDNA; 1230 BP.
XX
AC ABK86462;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human Apo-dystrophin-4 full length cDNA.
XX
KW Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	1. .1230
FT		/*tag= a
FT		/product= "Apo-dystrophin-4"
FT		/partial
FT		/note= "No start or stop codon shown"
FT		/transl_except= (pos:142. .144,aa:Xaa)
FT		/transl_except= (pos:241. .244,aa:Xaa)
FT		/transl_except= (pos:282. .284,aa:Xaa)
FT		/transl_except= (pos:421. .423,aa:Xaa)
FT		/transl_except= (pos:436. .438,aa:Xaa)

```
FT                   /transl_except= (pos:451. .453,aa:Xaa)
FT                   /transl_except= (pos:487. .489,aa:Xaa)
FT                   /transl_except= (pos:499. .504,aa:Xaa-Xaa)
FT                   /transl_except= (pos:628. .630,aa:Xaa)
FT                   /transl_except= (pos:781. .783,aa:Xaa)
FT                   /transl_except= (pos:799. .801,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:883. .885,aa:Xaa)
FT                   /transl_except= (pos:916. .920,aa:Xaa)
FT                   /transl_except= (pos:945. .947,aa:Xaa)
FT                   /transl_except= (pos:1024. .1026,aa:Xaa)
FT                   /transl_except= (pos:1075. .1077,aa:Xaa)
FT                   /transl_except= (pos:1084. .1086,aa:Xaa)
FT                   /transl_except= (pos:1114. .1116,aa:Xaa)
FT                   /transl_except= (pos:1141. .1143,aa:Xaa)
FT                   /transl_except= (pos:1186. .1188,aa:Xaa)
FT                   /transl_except= (pos:1216. .1218,aa:Xaa)
FT                   /transl_except= (pos:1222. .1224,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98729.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 4; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the full length cDNA sequence for human apo-
CC   dystrophin-4 containing a plurality of stop codons some of which may be
CC   read through due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 1230 BP; 404 A; 189 C; 259 G; 378 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 996;  DB 6;  Length 1230;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-211;
  Matches  996;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 294

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 354

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 414

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 474

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 534

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 594

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 654

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 714

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 774

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 834

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTTGTGAAGGGTAGTG 894

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 954

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 1014

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 1074

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 1134

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 1194

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db       1195 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 1230


RESULT 3
ABK86463
ID   ABK86463 standard; cDNA; 1234 BP.
XX
AC   ABK86463;
XX
```

```
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 full cDNA fragment #1.
XX
KW   Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 9; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is a human apo-dystrophin-4 fragment matching a region
CC   of mouse genomic DNA in the Apo-dystrophin-4 gene region
XX
SQ   Sequence 1234 BP; 405 A; 190 C; 260 G; 379 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 996;  DB 6;  Length 1234;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-211;
  Matches  996;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 298

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 358

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 418

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 478

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 538

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 598

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 658

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 718

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 778

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 838

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 898

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 958

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 1018

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 1078

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 1138

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 1198

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db       1199 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 1234

RESULT 4
ABK86464
ID   ABK86464 standard; DNA; 1044 BP.
XX
AC   ABK86464;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Mouse genomic sequence in the Apo-dystrophin-4 region, Mgen1073.
XX
KW   Mouse; ds; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Mus sp.
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
```

PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
XX
PT New human regulatory polynucleotide, useful for treating disorders
PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Disclosure; Fig 9; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is a region of mouse genomic DNA in the Apo-dystrophin-4
CC gene region
XX
SQ Sequence 1044 BP; 347 A; 177 C; 200 G; 320 T; 0 U; 0 Other;

Query Match 43.7%; Score 435; DB 6; Length 1044;
Best Local Similarity 84.7%; Pred. No. 7.1e-87;
Matches 577; Conservative 0; Mismatches 80; Indels 24; Gaps 7;

```
Qy        192 TTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTAACAAACAACGT 251
              || | ||    |||   ||  |  ||||||||  ||||||||  ||||||||     ||
Db        257 TTCTCCCCATCAAATGACACCATGCTGATCCAGTATTAAGCTAATACTAAC-----ACCA 311

Qy        252 GTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAAAATCAAACCTT 311
                |  ||||||||||| |||| ||||||||||  || ||||  ||| ||||  |  ||  |
Db        312 TGCAATGCTTCATTAACAAGGATTTGCTTCTTGCTAGAAATGGGTAAAAA--CGGACTGT 369

Qy        312 GTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATGGGGAATGATTT 371
              | | |||| ||| || |||||||||  ||||||||||   ||     ||  || ||||| |
Db        370 GGTCTGTATACCTTCAATGCAGCTTATGTGTTGTCTTTTCC--TGAAATGGTAATGACTC 427

Qy        372 CCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTAAAGTCTGTCTT 431
              || | |  |  ||  |  |  || ||       |||||  ||||||||      ||||
Db        428 CCAATAGTGGCAACCAGGGGTACAATACT-------TGCACACTTTGTAAA----CTCTT 476

Qy        432 TCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTT 491
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        477 TCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTT 536

Qy        492 GGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAAT 551
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        537 GGGCAGAGCGATGGAGTCCTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAAT 596

Qy        552 GTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAG 611
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        597 GTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAG 656

Qy        612 ACAGTAAGAGTTTACAAGAAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACT 670
              |||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||
Db        657 ACAGTAAGAGTTTACAAGAAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACT 716

Qy        671 GTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 GTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACG 776

Qy        731 TCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAAT 790
              |||||||||||||||||||||||| ||||||   ||||||||||||||||||||||||||
Db        777 TCTATGCAATTGTACAAAAAAGTTAAAAGAAA---ACATGTAAAATCTTGATAGCTAAAT 833

Qy        791 AACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTAT 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 AACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTAT 893

Qy        851 AAAGAAAGAATTATAAAGGAA 871
              |||||||| |   ||   ||
Db        894 AAAGAAAGATTGTAAACTAAA 914
```

RESULT 5
ABT33375
ID ABT33375 standard; DNA; 2463 BP.
XX
AC ABT33375;
XX
DT 22-MAY-2003 (first entry)
XX
DE NOVX DNA sequence SEQ ID No 64.
XX
KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
KW immunogen; non-human transgenic animal; gene therapy; gene; ds.
XX
OS Unidentified.
XX
PN WO200281517-A2.
XX
PD 17-OCT-2002.
XX
PF 22-JAN-2002; 2002WO-US002064.
XX
PR 19-JAN-2001; 2001US-0262892P.
PR 23-JAN-2001; 2001US-0263598P.
PR 24-JAN-2001; 2001US-0263799P.
PR 25-JAN-2001; 2001US-0264117P.
PR 25-JAN-2001; 2001US-0264139P.
PR 26-JAN-2001; 2001US-0264478P.
PR 30-JAN-2001; 2001US-0263351P.
PR 02-MAR-2001; 2001US-0272870P.
PR 14-MAR-2001; 2001US-0275927P.
PR 14-MAR-2001; 2001US-0275990P.
PR 15-MAR-2001; 2001US-0276449P.
PR 20-MAR-2001; 2001US-0277358P.
PR 23-MAR-2001; 2001US-0278151P.
PR 29-MAR-2001; 2001US-0279857P.
PR 20-APR-2001; 2001US-0285140P.
PR 20-APR-2001; 2001US-0285141P.
PR 30-APR-2001; 2001US-0287484P.
PR 17-MAY-2001; 2001US-0291701P.
PR 08-JUN-2001; 2001US-0296960P.
PR 10-JUL-2001; 2001US-0304353P.
PR 10-JUL-2001; 2001US-0304355P.
PR 12-JUL-2001; 2001US-0304886P.

PR 09-AUG-2001; 2001US-0311289P.
PR 13-AUG-2001; 2001US-0311975P.
PR 16-AUG-2001; 2001US-0312937P.
PR 18-OCT-2001; 2001US-0330227P.
PR 29-NOV-2001; 2001US-0334198P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;
XX
DR WPI; 2003-058504/05.
DR P-PSDB; ABJ37909.
XX
PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.
XX
PS Claim 9; Page 171; 672pp; English.
XX
CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
CC acids can be used in gene therapy. This polynucleotide sequence
CC represents a NOVX DNA sequence of the invention
XX
SQ Sequence 2463 BP; 691 A; 599 C; 580 G; 593 T; 0 U; 0 Other;

Query Match 41.4%; Score 412.2; DB 7; Length 2463;
Best Local Similarity 97.0%; Pred. No. 9.8e-82;
Matches 420; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy       439 CTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 498
            | | |||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1998 CCTAGTTCAAGAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 2057

Qy       499 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2058 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 2117

Qy       559 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2118 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 2177

Qy       619 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 678
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2178 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 2237

Qy       679 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2238 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 2297

Qy       739 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCC 798
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2298 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCC 2357

Qy       799 ATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAG 858
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2358 ATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAG 2417

Qy       859 AATTATAAAGGAA 871
             | |  ||  ||
Db      2418 ATTGTAAACTAAA 2430
```

RESULT 6
ABK81998
ID ABK81998 standard; DNA; 5339 BP.
XX
AC ABK81998;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR2-R21.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Example 6; Fig 13; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition

CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5339 BP; 1638 A; 1191 C; 1187 G; 1323 T; 0 U; 0 Other;

Query Match 41.2%; Score 410.8; DB 6; Length 5339;
Best Local Similarity 98.3%; Pred. No. 2.4e-81;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4551 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4610

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4611 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4730

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4731 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4790

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4791 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4850

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4851 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 4910

Qy       870 AA 871
            ||
Db      4911 AA 4912
```

RESULT 7
ABK81997
ID ABK81997 standard; DNA; 5417 BP.
XX
AC ABK81997;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR4-R23.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 12; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5417 BP; 1700 A; 1192 C; 1182 G; 1343 T; 0 U; 0 Other;

Query Match 41.2%; Score 410.8; DB 6; Length 5417;
Best Local Similarity 98.3%; Pred. No. 2.4e-81;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4628

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4629 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4688

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4689 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4748

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4749 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4808

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4809 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4868

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4869 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4928

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4929 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 4988

Qy       870 AA 871
            ||
Db      4989 AA 4990
```

RESULT 8
ABK81999
ID ABK81999 standard; DNA; 5462 BP.
XX
AC ABK81999;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR2-R21+H3.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.

XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 14; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5462 BP; 1668 A; 1225 C; 1212 G; 1357 T; 0 U; 0 Other;

Query Match 41.2%; Score 410.8; DB 6; Length 5462;
Best Local Similarity 98.3%; Pred. No. 2.4e-81;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4673

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4733

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4734 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4793

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4794 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4853

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4854 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4913

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4914 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4973

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
            ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||  ||
Db      4974 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 5033

Qy       870 AA 871
             ||
Db      5034 AA 5035
```

RESULT 9
ABK82000
ID ABK82000 standard; DNA; 8689 BP.
XX
AC ABK82000;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaH2-R19.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 15; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;

Query Match 41.2%; Score 410.8; DB 6; Length 8689;
Best Local Similarity 98.3%; Pred. No. 2.7e-81;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 6044

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 6104

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6105 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 6164

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6165 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 6224

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6225 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 6284

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6285 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 6344
```

```
Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6345 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 6404

Qy        870 AA 871
              ||
Db       6405 AA 6406
```

RESULT 10
ABK82002
ID ABK82002 standard; DNA; 11443 BP.
XX
AC ABK82002;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR9-R16.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 17; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;

  Query Match 41.2%; Score 410.8; DB 6; Length 11443;
  Best Local Similarity 98.3%; Pred. No. 2.8e-81;
  Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8739 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 8798

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 8858

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8859 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 8918

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8919 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 8978

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8979 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 9038

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9039 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 9098

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9099 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 9158

Qy        870 AA 871
              ||
Db       9159 AA 9160
```

RESULT 11
AAN90338
ID AAN90338 standard; cDNA; 12923 BP.
XX
AC AAN90338;
XX
DT 29-MAR-1992 (first entry)
XX
DE Sequence of human muscular dystrophy (MD) cDNA.
XX
KW Dystrophin; muscular dystrophy; probe; antibody; diagnosis; prenatal;
KW heterozygote; gene therapy; genetic screening; foetal screening; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 209. .12923
FT /*tag= a
XX
PN WO8906286-A.
XX
PD 13-JUL-1989.
XX
PF 16-DEC-1988; 88WO-US004504.
XX
PR 22-DEC-1987; 87US-00136618.
XX
PA (CHIL-) CHILDRENS MED CENT.
XX
PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;
XX
DR WPI; 1989-220587/30.
DR P-PSDB; AAP90373.
XX
PT Muscular dystrophy gene - used for prepn. of probes, dystrophic
PT polypeptide and antibodies for diagnosis and therapy of muscular
PT dystrophy.
XX
PS Disclosure; Fig 5; 68pp; English.
XX
CC The inventors claim an MD probe comprising a purified ss NA SQ which
CC hybridises to at least a part of the MD gene; pure dystrophin (DS)
CC polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
CC probes are equal to or greater than 10b of one of 12 cDNA sequences
CC deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd gene
XX
SQ Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T; 0 U; 7 Other;

  Query Match 41.2%; Score 410.8; DB 1; Length 12923;
  Best Local Similarity 98.3%; Pred. No. 2.9e-81;
  Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

RESULT 12
ABK81959
ID ABK81959 standard; DNA; 13957 BP.
XX
AC ABK81959;
XX
DT 13-AUG-2002 (first entry)
XX
DE cDNA encoding human dystrophin.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Example 2; Fig 1; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a human dystrophin polynucleotide
CC sequence used in the creation of the mini-dystrophin peptides of the
CC invention
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

```
  Query Match            41.2%;  Score 410.8;  DB 6;  Length 13957;
  Best Local Similarity  98.3%;  Pred. No. 3e-81;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

RESULT 13
ABT10904
ID ABT10904 standard; cDNA; 13957 BP.
XX
AC ABT10904;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 1038.
XX
KW Human; breast specific gene; breast cancer; differential expression;
KW cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US002176.
XX
PR 25-JAN-2001; 2001US-0263757P.
PR 25-APR-2001; 2001US-0286090P.
PR 23-MAY-2001; 2001US-0292517P.
XX
PA (GENE-) GENE LOGIC INC.

```
XX
PI   Orr MS,  Nation M,  Diggans JC,  Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 1038; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             41.2%;  Score 410.8;  DB 6;  Length 13957;
  Best Local Similarity   98.3%;  Pred. No. 3e-81;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |    ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674


RESULT 14
ABN95786
ID   ABN95786 standard; DNA; 13957 BP.
XX
AC   ABN95786;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2284 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US030589.
XX
PR   02-OCT-2000; 2000US-0237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer, hepatocellular
PT   carcinoma or metastatic liver tumor in a patient, involves detecting the
PT   level of expression of two or more genes in a liver tissue sample.
XX
PS   Claim 1; SEQ ID NO 2284; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             41.2%;  Score 410.8;  DB 6;  Length 13957;
  Best Local Similarity   98.3%;  Pred. No. 3e-81;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

RESULT 15
ABS69900
ID ABS69900 standard; DNA; 13957 BP.
XX
AC ABS69900;
XX
DT 21-NOV-2002 (first entry)
XX
DE Human dystrophin gene.
XX
KW Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;
KW adenosine deaminase deficiency; severe combined immune deficiency; PAH;
KW beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;
KW low density lipoprotein gene; familial hypercholesterolaemia;
KW hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;
KW phenylalanine hydroxylase gene; gene therapy; phenylketonuria;
KW dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;
KW human cystic fibrosis transmembrane conductance regulator gene;
KW antianaemic; antilipaemic; nootropic; cytostatic; dermatological; human;
KW alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;
KW lysosomal arylsulfatase A; omithine transcarbamylase; ARSA; OTC; NP;
KW purin nucleoside phosphorylase; gene; ds.
XX
OS Homo sapiens.
XX
PN US2002102731-A1.
XX
PD 01-AUG-2002.
XX
PF 12-FEB-2001; 2001US-00782378.
XX
PR 02-OCT-2000; 2000US-0237747P.
XX
PA (UYNY) UNIV NEW YORK STATE RES FOUND.
XX
PI Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;
XX
DR WPI; 2002-690619/74.
XX
PT Producing vector, by introducing vector having nucleotide sequence,
PT adenovirus inverted terminal repeats and packaging sequence, and adeno-
PT associated virus terminal repeat, into cell, and culturing cell.
XX
PS Disclosure; Page 122-128; 191pp; English.
XX
CC The present invention relates to a new method of producing a vector. The
CC method involves introducing recombinant vector having nucleotide sequence
CC (NS) having 5' and 3' end, left and right inverted terminal repeats of
CC adenovirus flanking NS, adenovirus packaging sequence linked to inverted
CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'
CC end of NS, into cell expressing adenovirus early gene lacking from vector
CC ; and culturing cell to produce another vector. The method is useful for
CC generating vectors, especially mAd vectors. The method is useful in
CC transferring nucleotide sequences of interest into a cell, for gene
CC transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.
CC The nucleotide sequences are useful for treating diseases associated with
CC it, i.e. adenosine deaminase gene associated with adenosine deaminase
CC deficiency with severe combined immune deficiency, beta-chain of
CC haemoglobin gene associated with beta-thalassaemia and sickle cell
CC disease, receptor for low density lipoprotein gene associated with
CC familial hypercholesterolaemia, hypoxanthine-guanine
CC phosphoribosyltransferase associated with Lesch-Nyhan syndrome,
CC phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,
CC dystrophin gene associated with muscular dystrophy, and human cystic
CC fibrosis transmembrane conductance regulator gene associated with cystic
CC fibrosis. The present nucleic acid sequence represents a human disease
CC gene sequence that was used in the methods of the invention
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 41.2%; Score 410.8; DB 6; Length 13957;
Best Local Similarity 98.3%; Pred. No. 3e-81;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

Search completed: April 6, 2004, 11:42:39
Job time : 473.188 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 08:36:01 ; Search time 110.764 Seconds
(without alignments)
4990.154 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	409.2	41.1	13977	4	US-09-484-970B-60	Sequence 60, Appl
c 2	374.2	37.6	19307	3	US-08-836-022A-10	Sequence 10, Appl
c 3	374.2	37.6	19307	3	US-09-427-048A-10	Sequence 10, Appl
c 4	140.4	14.1	13977	4	US-09-484-970B-60	Sequence 60, Appl
5	101.4	10.2	19307	3	US-08-836-022A-10	Sequence 10, Appl
6	101.4	10.2	19307	3	US-09-427-048A-10	Sequence 10, Appl
c 7	45	4.5	615	3	US-08-998-416-186	Sequence 186, App
8	44	4.4	8093	4	US-10-204-708-31	Sequence 31, Appl
c 9	43.6	4.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
10	43	4.3	1830121	4	US-09-557-884-1	Sequence 1, Appli
11	43	4.3	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c 12	42.8	4.3	7218	1	US-08-232-463-14	Sequence 14, Appl
13	42.4	4.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
14	42	4.2	5156	2	US-09-091-432-3	Sequence 3, Appli
15	42	4.2	5156	4	US-09-387-663-3	Sequence 3, Appli
16	42	4.2	5156	4	US-09-214-139B-3	Sequence 3, Appli
c 17	41.6	4.2	640681	4	US-09-790-988-1	Sequence 1, Appli
18	41.2	4.1	11049	4	US-10-204-708-21	Sequence 21, Appl
19	41	4.1	15567	4	US-09-627-376-3	Sequence 3, Appli
20	40.6	4.1	615	3	US-08-998-416-186	Sequence 186, App
21	40.6	4.1	636	3	US-08-998-416-1137	Sequence 1137, Ap
22	40.6	4.1	837	3	US-08-998-416-288	Sequence 288, App
c 23	40.4	4.1	5152	4	US-10-204-708-73	Sequence 73, Appl
c 24	40.4	4.1	6124	4	US-08-213-419B-3	Sequence 3, Appli
c 25	40.4	4.1	319608	4	US-09-539-333D-1	Sequence 1, Appli
c 26	40.4	4.1	319608	4	US-09-679-409-1	Sequence 1, Appli
27	40	4.0	6124	4	US-08-213-419B-3	Sequence 3, Appli
28	39.8	4.0	4185	4	US-09-417-485D-7	Sequence 7, Appli
29	39.8	4.0	8920	2	US-08-446-855A-1	Sequence 1, Appli
30	39.8	4.0	8920	3	US-09-150-741-1	Sequence 1, Appli
31	39.8	4.0	10640	4	US-09-417-485D-5	Sequence 5, Appli
c 32	39.6	4.0	17310	4	US-08-956-171E-23	Sequence 23, Appl
33	39.4	4.0	1661	3	US-08-821-994-82	Sequence 82, Appl
34	39.4	4.0	19233	4	US-10-204-708-46	Sequence 46, Appl
35	39.2	3.9	640681	4	US-09-790-988-1	Sequence 1, Appli
36	39	3.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
37	38.8	3.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
38	38.8	3.9	51952	3	US-08-947-823-1	Sequence 1, Appli
39	38.4	3.9	1121	6	5256558-7	Patent No. 5256558
40	38.4	3.9	1241	1	US-07-593-657-6	Sequence 6, Appli
41	38.4	3.9	1241	3	US-08-942-012B-3	Sequence 3, Appli
c 42	38.4	3.9	3312	4	US-09-601-198-58	Sequence 58, Appl
c 43	38.4	3.9	4600	4	US-09-702-705-1797	Sequence 1797, Ap
c 44	38.4	3.9	4600	4	US-09-736-457-1797	Sequence 1797, Ap
c 45	38.4	3.9	4600	4	US-09-671-325-1797	Sequence 1797, Ap

ALIGNMENTS

RESULT 1
US-09-484-970B-60
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
; NAME/KEY: unsure
; LOCATION: 11721-11761, 12294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match 41.1%; Score 409.2; DB 4; Length 13977;
Best Local Similarity 98.1%; Pred. No. 1.6e-92;
Matches 414; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11254 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11313

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11314 CTTAGTATCATTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11373

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11374 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11433

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11434 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11493

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11494 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11553
```

```
Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11554 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11613

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              |||||||||||||||||||||||||||||||||||||||||||||||||||||  |   ||
Db      11614 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11673

Qy        870 AA 871
              ||
Db      11674 AA 11675
```

RESULT 2
US-08-836-022A-10/c
; Sequence 10, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Chen, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,022A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/331,381
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19307 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-836-022A-10

Query Match 37.6%; Score 374.2; DB 3; Length 19307;
Best Local Similarity 96.0%; Pred. No. 9.5e-84;
Matches 406; Conservative 0; Mismatches 13; Indels 4; Gaps 2;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       3441 AGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 3382

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3381 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 3322

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       3321 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 3262

Qy        630 AAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       3261 AAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTC 3202

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 748
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3201 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 3142

Qy        749 AAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 808
              ||||||| ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       3141 AAAGTTAAAAGAAA---ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 3085

Qy        809 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAG 868
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||| |   ||
Db       3084 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAGAGATTGTAAACT 3025

Qy        869 GAA 871
              |||
Db       3024 AAA 3022
```

RESULT 3
US-09-427-048A-10/c
; Sequence 10, Application US/09427048A
; Patent No. 6203975
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; Wilson, James M.
; Fisher, Krishna J.
; Chen, Shu-Jen
; Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; Methods of Use Thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/427,048A
; FILING DATE: 21-Oct-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,022
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19307 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

```
  Query Match            37.6%;  Score 374.2;  DB 3;  Length 19307;
  Best Local Similarity  96.0%;  Pred. No. 9.5e-84;
  Matches  406;  Conservative    0;  Mismatches   13;  Indels    4;  Gaps    2;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       3441 AGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 3382

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3381 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 3322

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       3321 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 3262

Qy        630 AAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       3261 AAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTC 3202

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 748
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3201 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 3142

Qy        749 AAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 808
              ||||||| ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       3141 AAAGTTAAAAGAAA---ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 3085

Qy        809 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAG 868
              |||||||||||||||||||||||||||||||||||||||||||| ||| |   ||
Db       3084 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAGAGATTGTAAACT 3025

Qy        869 GAA 871
              |||
Db       3024 AAA 3022
```

```
RESULT 4
US-09-484-970B-60/c
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Karen A.
;  APPLICANT:  Volkmuth, Wayne
;  APPLICANT:  Walker, Michael G.
;  TITLE OF INVENTION:  BONE REMODELING GENES
;  FILE REFERENCE:  PB-0014 US
;  CURRENT APPLICATION NUMBER: US/09/484,970B
;  CURRENT FILING DATE:  2000-01-18
;  NUMBER OF SEQ ID NOS:  172
;  SOFTWARE:  PERL Program
; SEQ ID NO 60
;   LENGTH: 13977
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
;   NAME/KEY: unsure
;   LOCATION: 11721-11761, 12294, 13969
;   OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

  Query Match            14.1%;  Score 140.4;  DB 4;  Length 13977;
  Best Local Similarity  90.4%;  Pred. No. 1.8e-25;
  Matches  161;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy        820 TTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAA 879
              |||||||| | |    |  ||  |   | |||||||||||||||||||||||||||||||
Db      13346 TTTGGGTTTTCTTTTGAAAATTATGAAGGAAAAAGAAAGAATTATAAAGGAAAAAGAAAA 13287

Qy        880 TAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACC 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13286 TAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACC 13227

Qy        940 CCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAG-CCGTGTTTGATGTTAATTAATT 996
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      13226 CCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCCGTGTTTGATGTTAATTAATT 13169
```

```
RESULT 5
US-08-836-022A-10
; Sequence 10, Application US/08836022A
; Patent No. 6001557
;  GENERAL INFORMATION:
;    APPLICANT:  Trustees of the University of Pennsylvania
;    APPLICANT:  Wilson, James M.
;    APPLICANT:  Fisher, Krishna J.
;    APPLICANT:  Chen, Shu-Jen
;    APPLICANT:  Weitzman, Matthew
;    TITLE OF INVENTION:  Improved Adenovirus Virus and
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P O Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,022A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/331,381
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GNVPN.008PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19307 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-836-022A-10

  Query Match            10.2%;  Score 101.4;  DB 3;  Length 19307;
  Best Local Similarity  81.8%;  Pred. No. 1e-15;
  Matches  117;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy        851 AAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA 910
              ||  | | |   | ||||||||  |    ||  |||  |||||||||||||| ||||||
Db       1480 AATTATAAATGAAAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAA 1539

Qy        911 CTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 970
              | ||||||||||||||  || ||||||||||||||||||||||||||||||||||||||
Db       1540 ATTAGGTGTGCACAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 1599

Qy        971 GAAGCCGTGTTTGATGTTAATTA 993
              ||||| ||||||||||||||||||
Db       1600 GAAGCTGTGTTTGATGTTAATTA 1622
```

```
RESULT 6
US-09-427-048A-10
```

```
; Sequence 10, Application US/09427048A
; Patent No. 6203975
;  GENERAL INFORMATION:
;       APPLICANT: Trustees of the University of Pennsylvania
;                  Wilson, James M.
;                  Fisher, Krishna J.
;                  Chen, Shu-Jen
;                  Weitzman, Matthew
;       TITLE OF INVENTION: Improved Adenovirus Virus and
;                           Methods of Use Thereof
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Cntr, P O Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/427,048A
;            FILING DATE: 21-Oct-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/836,022
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;  INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 19307 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

  Query Match             10.2%;  Score 101.4;  DB 3;  Length 19307;
  Best Local Similarity   81.8%;  Pred. No. 1e-15;
  Matches  117;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy        851 AAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA 910
              ||  | | |   | ||||||| |     || ||| ||||||||||||||| ||||||
Db       1480 AATTATAAATGAAAAAAGGAAACTGGTGTCCACACAACAGACAAGTGGTGAAGTTGTGAA 1539

Qy        911 CTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 970
              | ||||||||||||||||  || ||||||||||||||||||||||||||||||||||||
Db       1540 ATTAGGTGTGCACAATTACTAGAAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 1599

Qy        971 GAAGCCGTGTTTGATGTTAATTA 993
              ||||| ||||||||||||||||||
Db       1600 GAAGCTGTGTTTGATGTTAATTA 1622


RESULT 7
US-08-998-416-186/c
; Sequence 186, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  186:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1074RP
US-08-998-416-186

  Query Match             4.5%;  Score 45;  DB 3;  Length 615;
  Best Local Similarity   48.0%;  Pred. No. 0.037;
  Matches  129;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        583 TATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATA 642
              || ||||||| ||  || ||  |  | |   | |||     || | |||| ||| ||
Db        576 TAAAATATTAATTAAATAATTAATTTAATTAAATAAATAAATAAATAAAATGAATATAAT 517

Qy        643 TTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTG 702
              || ||    |    ||      |      |   | ||| |||       ||| | |  || |
Db        516 TTATTAAATAATATAAAAATTAATAATAAGAAATTAAAGTTAAAATTAATTTTAATAATA 457

Qy        703 TTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAA 762
              || | | || |  | | || | || |  |||  |  |||||  |||  | |
Db        456 ATTCTTATAAAAAGTTAAATAATATAAATCAACATAATATTTATAAAAATAGATATTATA 397

Qy        763 ACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTT 822
              |  | |   | ||| |  |   | ||  |  |   || | | ||    ||   || ||
Db        396 ATAAAAATAATAATTTACAATATTTAAATAAATTAAATCTTTATAATAATAAAATATATT 337

Qy        823 GGGTTGTTTAAAAATTTATAACAGTTATA 851
              ||  | || |||| ||||    ||||
Db        336 ATTTTAATAAACAATTAATAAATAATATA 308


RESULT 8
US-10-204-708-31
; Sequence 31, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
```

```
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 31
;   LENGTH: 8093
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-31

  Query Match             4.4%;  Score 44;  DB 4;  Length 8093;
  Best Local Similarity   48.7%;  Pred. No. 0.16;
  Matches  149;  Conservative    0;  Mismatches  155;  Indels    2;  Gaps    1;

Qy        579 TTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATC 638
              |||| || ||||| ||      || |||||||  |  |  |  ||     ||| |||
Db        336 TTTAATTATATTTATTGTTATAATAGGATTAGTAATAATTATATTTATTTAAAAAAAAA 395

Qy        639 TATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTT 698
              | ||| || || ||         |  | ||| | ||||    ||||   | ||  | | ||
Db        396 AAGATTGTTATGTAGTTATAGAATTATTTAATATAGAAAATAGTAAAATATATTTTTTTT 455

Qy        699 ATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAA 758
              || |||| ||      |      || || |      |     |     |||    |||||
Db        456 TTTTTTTTTTTTTTTTTTTTTTTATAAATAAGATTAGTTTATAGTAAATTTTTTATAG 515

Qy        759 GAAAAC--TACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACG 816
              ||    ||  |  ||| | ||| |  ||  |  ||    ||  |  |||   | |
Db        516 TTAAGGGTTAATTTAAAATTTTGGTTTATATTTTTTTTTTATTTAATGATTATATGATG 575

Qy        817 CATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGA 876
              |  ||   ||  |||| | ||||    |  ||||  | ||   |   |||  | | ||||
Db        576 TAAATTAATTTTATTAATATTTTAAGTAAAATATATTGGAATTTTATAAAATGTATAAGA 635

Qy        877 AAATAA 882
                |||
Db        636 TTTTAA 641


RESULT 9
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             4.4%;  Score 43.6;  DB 4;  Length 832;
  Best Local Similarity   16.0%;  Pred. No. 0.092;
  Matches   67;  Conservative  165;  Mismatches  184;  Indels    2;  Gaps    1;

Qy        576 TGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAA 635
              ||  ||| |  ||    |  ||  || |   | | |   |  |  | | ||| | ||||||
Db        433 TGACTTTGACTATGAACTAACTCCAGATTTGCTAAAGTACAAGAATTGTACCATAAATAA 374

Qy        636 ATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCT 695
              | ||| || |: : :::  :  |  |:| :::::::|:::::  ::       :  ::: :
Db        373 TTATATTTTGTYYWTWWKTWWYWTTYTWTRMWWMKKKARRWYYWWKSTYACASRYRKYTW 314

Qy        696 G--TTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGT 753
              |  ::::::   |:::  : ::  : :|    :: :    ::::|: : : :: |: : |
Db        313 GWWWYMWKRMMSTRWYCYMCWKCCMYRGRRCAWYTMARGRMWSYAWGKWKSMRSAMSMCT 254

Qy        754 TATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGA 813
               ::  :     |: |   :  :::: :::      :  :  |::  :|:|    ::
Db        253 RMYYKKGSTYWTMKCTCATWCYWYWKYWKRMWSKTCWSGSRGGYMTSYTSTRSYSMYWAS 194

Qy        814 ACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAAAA 873
              ::  : :::|: :  ::::|:  ::::|   ::  : ::: |:  ::: ::|:  : : |
Db        193 WMYTMCWWWGRWWSTYWYMAWGKKWWRYATTWRRAMMWWWAAWTMMWYMWWAWCMSSRGA 134

Qy        874 AGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGG 933
              | ::  :: |:: : :    :  :  :|  ::|:: ::   :::::  :::  : :| |
Db        133 AMYRRTMMMWGYRYWWRKKSYRRTRCAWAYAWKTKRSYYWCWRWKWKRCMMMMMAMAYG 74

Qy        934 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGCCGTGTTTGATGTTAAT 991
              :::  |:  :: ::: : : : :  :: : ::: : : :   :::: ::   ::|:
Db         73 KTMMMRACWKTRYWRWWAWAMWRMWWTMMMMYYWYWRAMKRRWMWRKWRSWSWMWMAW 16


RESULT 10
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/557,884
;            FILING DATE: 25-Apr-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/476,102
;            FILING DATE: JUN-5-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P3
;       TELECOMMUNICATION INFORMATION:
```

```
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match              4.3%;  Score 43;  DB 4;  Length 1830121;
  Best Local Similarity   48.7%;  Pred. No. 1.9;
  Matches  115;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        738 AATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGC 797
              || |||   ||  | | |||| || |       | | ||   ||||| ||||| ||
Db     413354 AAGTGTTATAACCAATAATAAAAAGAGATGGAATTAGATAAAGATAGATAAATGCCTGTT 413413

Qy        798 CATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAA 857
              |    || || | ||     || | |||||| | ||  | ||  |   |   | ||||||
Db     413414 TAAAGTTTAATTTTGATTTTCTTATAGGTTGCTCAATTAAATAATAACTTGTGATAGAAA 413473

Qy        858 GAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGT 917
              ||| |||     | | |  ||  |     ||  | | |  |||| |||    ||
Db     413474 AAATAATAGTTAATATAACAACAATTAAAACTGTAATGCCTTGAATCTGCTTGCTGCCAG 413533

Qy        918 GTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAA 973
              ||| | ||    ||    ||| |||| | |   | ||  || || || | |
Db     413534 TAATGTAATAAGCAAACGCAATAAAAATCCAATGATATAAATAANGAGAATAAGGA 413589


RESULT 11
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;  GENERAL INFORMATION:
;       APPLICANT: Robert D. Fleischmann
;                  Mark D. Adams
;                  Owen White
;                  Hamilton O. Smith
;                  J. Craig Venter
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville,
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/643,990A
;            FILING DATE: 23-Aug-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/487,429
;            FILING DATE: 1995-06-07
;            APPLICATION NUMBER: 08/426,787
;            FILING DATE: 1995-04-21
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kenley K. Hoover
;            REGISTRATION NUMBER: 40,302
;            REFERENCE/DOCKET NUMBER: PB186P1C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match              4.3%;  Score 43;  DB 4;  Length 1830121;
  Best Local Similarity   48.7%;  Pred. No. 1.9;
  Matches  115;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        738 AATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGC 797
              || |||   ||  | | |||| || |       | | ||   ||||| ||||| ||
Db     413354 AAGTGTTATAACCAATAATAAAAAGAGATGGAATTAGATAAAGATAGATAAATGCCTGTT 413413

Qy        798 CATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAA 857
              |    || || | ||     || | |||||| | ||  | ||  |   |   | ||||||
Db     413414 TAAAGTTTAATTTTGATTTTCTTATAGGTTGCTCAATTAAATAATAACTTGTGATAGAAA 413473

Qy        858 GAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGT 917
              ||| |||     | | |  ||  |     ||  | | |  |||| |||    ||
Db     413474 AAATAATAGTTAATATAACAACAATTAAAACTGTAATGCCTTGAATCTGCTTGCTGCCAG 413533

Qy        918 GTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAA 973
              ||| | ||    ||    ||| |||| | |   | ||  || || || | |
Db     413534 TAATGTAATAAGCAAACGCAATAAAAATCCAATGATATAAATAANGAGAATAAGGA 413589


RESULT 12
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 4.3%; Score 42.8; DB 1; Length 7218;
Best Local Similarity 15.0%; Pred. No. 0.31;
Matches 41; Conservative 118; Mismatches 115; Indels 0; Gaps 0;

Qy 701 TGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGA 760
Db 1518 TATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAAGTAGT 1459

Qy 761 AAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATT 820
Db 1458 TAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1399

Qy 821 TTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAAT 880
Db 1398 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1339

Qy 881 AACGCAATGGACAAGTGGTGAAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCC 940
Db 1338 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1279

Qy 941 CAAAACCAAAGTGAGGTAGAAATAGCATGAGAAG 974
Db 1278 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1245

RESULT 13
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 4.3%; Score 42.4; DB 2; Length 19124;
Best Local Similarity 50.8%; Pred. No. 0.54;
Matches 127; Conservative 0; Mismatches 121; Indels 2; Gaps 1;

Qy 640 ATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTA 699
Db 4746 ATAATATATTCATATGTAGTCATAGTGTCAATGAATATAAAATATGGTATATTTATATTA 4805

Qy 700 TTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAG 759
Db 4806 TTGTATATATTAAATAAGTAACACAGAACATTATATATAGTAATAAATAGAAGAAATAAT 4865

Qy 760 AAAACTACATGTAAAATCTTGATAGCTA--AATAACTTGCCATTTCTTTATATGGAACGC 817
Db 4866 ATATTTTTATGTTATATATTATTAGTTATTATAAAGGGGAAAATTCATAATATTTATGAA 4925

Qy 818 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAA 877
Db 4926 AATTTTTGTATATGATATAGTTATAAGTTAAAAAAAAAAAAAAACAAGAACAAAAATGGA 4985

Qy 878 AATAACGCAA 887
Db 4986 AAGCATAAAA 4995

RESULT 14
US-09-091-432-3
; Sequence 3, Application US/09091432
; Patent No. 5981837
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
; FILE REFERENCE: 7024-325
; CURRENT APPLICATION NUMBER: US/09/091,432
; CURRENT FILING DATE: 1998-06-18
; EARLIER APPLICATION NUMBER: PCT/US96/20094
; EARLIER FILING DATE: 1996-12-19
; EARLIER APPLICATION NUMBER: US 60/009,119
; EARLIER FILING DATE: 1995-12-22
; EARLIER APPLICATION NUMBER: US 60/013,388
; EARLIER FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Microsoft Word 2.0C
; SEQ ID NO 3
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-091-432-3

Query Match 4.2%; Score 42; DB 2; Length 5156;
Best Local Similarity 48.7%; Pred. No. 0.44;
Matches 114; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 559 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 618
Db 1156 AAGTCTTGTGTAACGCACTGTATCTAAATTGTGAGTAACGACAAAATAATTAAAATTAAA 1215

Qy 619 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 678
Db 1216 GGACCTTCAAGTATTATTAGTATCTCTGTCTAAGATGCACAGGTATTCAGTAATAGTAAT 1275

Qy 679 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 738
Db 1276 AAATAATTACTTGTATAATTAATATCTAATTAGTAAACCTTGTGTCTAAACCTAAATGAG 1335

Qy 739 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAA 792

```
                |  |   |||||  | ||   ||    |  | |||   | ||  | |||  ||
Db       1336 CATAAATCCAAAAGCAAAAATCTAAACCTAACTGAAAAAGTCATTACGAAAAAA 1389
```

RESULT 15
US-09-387-663-3
; Sequence 3, Application US/09387663
; Patent No. 6489538
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
; FILE REFERENCE: 7024-325
; CURRENT APPLICATION NUMBER: US/09/387,663
; CURRENT FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: US 09/091,432
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: PCT/US96/20094
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: US 60/009,119
; PRIOR FILING DATE: 1995-12-22
; PRIOR APPLICATION NUMBER: US 60/013,388
; PRIOR FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Microsoft Word 2.0C
; SEQ ID NO 3
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-387-663-3

Query Match 4.2%; Score 42; DB 4; Length 5156;
Best Local Similarity 48.7%; Pred. No. 0.44;
Matches 114; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy        559 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 618
              || || ||  |  ||||  || | || | | |   ||     || |  |||| |    ||
Db       1156 AAGTCTTGTGTAACGCACTGTATCTAAATTGTGAGTAACGACAAAATAATTAAAATTAAA 1215

Qy        619 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 678
              |    | |||| | ||    ||| | | | | |   | |  ||||||      |||   |
Db       1216 GGACCTTCAAGTATTATTAGTATCTCTGTCTAAGATGCACAGGTATTCAGTAATAGTAAT 1275

Qy        679 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 738
              | || || ||      ||| | | |  |||  ||      |  | || || |  |||
Db       1276 AAATAATTACTTGTATAATTAATATCTAATTAGTAAACCTTGTGTCTAAACCTAAATGAG 1335

Qy        739 ATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAA 792
              |  |   |||||  | ||   ||    |  | |||   | ||  | ||| ||
Db       1336 CATAAATCCAAAAGCAAAAATCTAAACCTAACTGAAAAAGTCATTACGAAAAAA 1389
```

Search completed: April 6, 2004, 14:14:25
Job time : 118.764 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:49:50 ; Search time 3674.56 Seconds
(without alignments)
8094.223 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	459.2	46.1	525	28	AQ679243	AQ679243 HS_5417_A
2	443.2	44.5	472	28	AQ583263	AQ583263 RPCI-11-4
3	410.8	41.2	844	12	BG706268	BG706268 602669567
4	375.8	37.7	4437	11	AK036936	AK036936 Mus muscu
5	370.4	37.2	439	14	CB750007	CB750007 AMGNNUC:S
6	354	35.5	924	13	BQ927942	BQ927942 AGENCOURT
7	342.8	34.4	629	9	AL712236	AL712236 DKFZp686J
8	342.8	34.4	629	13	BX509925	BX509925 DKFZp686J
9	296	29.7	633	13	BU337294	BU337294 603515757
c 10	291.2	29.2	531	12	BI281078	BI281078 UI-R-DD0-
11	286.2	28.7	854	13	BU226117	BU226117 603947264
12	282.8	28.4	739	13	BU368789	BU368789 603596716
c 13	282.6	28.4	500	9	AI008807	AI008807 EST203258
c 14	281.2	28.2	576	12	BG795230	BG795230 UTSW_SM30
15	278.6	28.0	421	13	BY635064	BY635064 BY635064
16	273.8	27.5	658	13	BU352875	BU352875 603525781
c 17	271.6	27.3	308	14	N75050	N75050 za89c07.s1
18	270.8	27.2	838	13	BU352423	BU352423 603530318
19	269	27.0	440	10	BB819684	BB819684 BB819684
20	266.2	26.7	687	14	W05777	W05777 za89c07.r1
21	264.2	26.5	929	13	BU274687	BU274687 603531284
22	263.2	26.4	778	13	BU309070	BU309070 603538118
23	260.2	26.1	492	14	CB725360	CB725360 AMGNNUC:N
24	245.4	24.6	999	13	BX419180	BX419180 BX419180
25	245	24.6	648	13	BU402013	BU402013 604138944
26	244.8	24.6	739	9	AU138648	AU138648 AU138648
27	237.2	23.8	826	13	BU104673	BU104673 603004172
28	233.8	23.5	1201	13	BX463891	BX463891 BX463891
29	231.2	23.2	859	13	BU300671	BU300671 603734536
30	229.2	23.0	928	13	BU107395	BU107395 603110749
31	221.2	22.2	927	13	BU139194	BU139194 603135045
32	217.8	21.9	888	14	CF552038	CF552038 AGENCOURT
33	215	21.6	277	10	AW861957	AW861957 CM3-CT034
34	213.8	21.5	275	10	AW604756	AW604756 CM0-CT034
35	213.8	21.5	284	10	AW858457	AW858457 CM3-CT034
36	208	20.9	234	14	N86611	N86611 J9202F Huma
c 37	202.8	20.4	619	10	AW771158	AW771158 hn53c06.x
c 38	202.2	20.3	271	10	AW858578	AW858578 CM3-CT034
c 39	195.6	19.6	886	13	BX760461	BX760461 BX760461
c 40	195.4	19.6	851	13	BX725372	BX725372 BX725372
c 41	194.8	19.6	628	13	BX693509	BX693509 BX693509
42	193.8	19.5	677	13	BX855319	BX855319 BX855319
43	192.8	19.4	618	9	AL793265	AL793265 AL793265
c 44	191	19.2	537	12	BJ090827	BJ090827 BJ090827
45	186.6	18.7	351	13	BY673194	BY673194 BY673194

ALIGNMENTS

RESULT 1
AQ679243
LOCUS AQ679243 525 bp DNA linear GSS 25-JUN-1999
DEFINITION HS_5417_A1_B06_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=993 Col=11 Row=C, genomic survey sequence.
ACCESSION AQ679243
VERSION AQ679243.1 GI:5228047
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

```
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 993  row: C  column: 11
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=993 Col=11 Row=C"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             46.1%;  Score 459.2;  DB 28;  Length 525;
  Best Local Similarity   95.4%;  Pred. No. 3.4e-72;
  Matches  481;  Conservative    0;  Mismatches   22;  Indels    1;  Gaps    1;

Qy        137 TCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAATGTAGCTTGACATTTTT 196
              ||| | | | |     |||||||||| ||||||||||||||||||||||||||||||| ||
Db         23 TCTCTCGGACTGCCNTGATATATATNTACATTGGGAACCTGAATGTAGCTTGACATTNTT 82

Qy        197 CCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTAACAAACAACGTGTAAT 256
              |||||| ||||||| | |  ||||||||||||||||||||||||||||||||||||||
Db         83 CCATGTTAACACCATTTGTTTGATCCAACATTAAGCTGATACTAACAAACAACGTGTAAT 142

Qy        257 GGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAAAATCAAACCTTGTTGT 316
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAAAATCAAACCTTGTTGT 202

Qy        317 GTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATGGGGAATGATTTCCCAA 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATGGGGAATGATTTCCCAA 262

Qy        377 ATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTAAAGTCTGTCTTTCTTT 436
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        263 ATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTG-AAAGTCTGTCTTTCTTT 321

Qy        437 CTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCA 496
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 CTCTTTTGTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCA 381

Qy        497 GAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTT 556
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        382 GAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGGTTT 441

Qy        557 ACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGT 616
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        442 ACAACTCCTGATTCCCGCATGGTTNTTATAATATTCATACAACAAAGAGGATTAGACACG 501

Qy        617 AAGAGTTTACAAGAAATAAATCTA 640
              |||||||||||||||||||||||
Db        502 TAGAGTTTACAAGAAATAAATCTA 525


RESULT 2
AQ583263
LOCUS       AQ583263                472 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  RPCI-11-451G14.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-451G14, genomic survey sequence.
ACCESSION   AQ583263
VERSION     AQ583263.1  GI:5010373
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7672957"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-451G14"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             44.5%;  Score 443.2;  DB 28;  Length 472;
  Best Local Similarity   96.2%;  Pred. No. 2.5e-69;
  Matches  454;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy        172 AACCTGAATGTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAG 231
              |||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db          1 AACCTGAATGTAGCTAGACATTTTTCCATGTTAACACCAGTAGCCTGATCCAACATTTAG 60

Qy        232 CTGATACTAACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAA 291
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db         61 CTGATACTAACAAACAACGTGTAATGGCTTCATTAATAAGGCTTCGCTTCTTCCTGGAAA 120

Qy        292 CTGGTGAAAAATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCAC 351
              |||| |||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        121 CTGGCGAAAAATCAAACCTTGTTGGGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCAC 180

Qy        352 CCAGAAATGGGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCA 411
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGAAATGGCGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCA 240

Qy        412 CCTTTTGTAAAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTT 471
              |||||||||||||||| |||||||||||||||||||||| |||||||||| ||||||||
Db        241 CCTTTTGTAAAGTCTGACTTTCTTTCTCTTTGTTTTCCACGACACAATGTACGAAGTCTT 300

Qy        472 TTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGAT 531
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCCACATGGCATATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGAT 360

Qy        532 GAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATT 591
              |||||| |||||||||||||| |||||||| |||||||||||||||||||||| ||||||
Db        361 GAAGAAAGAGCAGAATAAATGGTTTACAACTGCTGATTCCCGCATGGTTTTTATTATATT 420
```

```
Qy        592 CATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 643
              ||||||||||||||||| ||||||| |||||||||||||| ||||||||||||
Db        421 CATACAACAAAGAGGATGAGACAGTTAGAGTTTACAAGACATAAATCTATAT 472


RESULT 3
BG706268
LOCUS       BG706268                844 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602669567F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4792313 5',
            mRNA sequence.
ACCESSION   BG706268
VERSION     BG706268.1  GI:13981445
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10670  row: e  column: 18
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4792313"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             41.2%;  Score 410.8;  DB 12;  Length 844;
  Best Local Similarity   98.3%;  Pred. No. 1.2e-63;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 147

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 207

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 267

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 327

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 387

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 447

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db        448 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 507

Qy        870 AA 871
              ||
Db        508 AA 509


RESULT 4
AK036936
LOCUS       AK036936                4437 bp    mRNA    linear   HTC 19-SEP-2003
DEFINITION  Mus musculus adult female vagina cDNA, RIKEN full-length enriched
            library, clone:9930028B14 product:dystrophin, muscular dystrophy,
            full insert sequence.
ACCESSION   AK036936
VERSION     AK036936.1  GI:26085504
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4437)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
```

Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.

FEATURES Location/Qualifiers

source 1. .4437
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9930028B14"
/db_xref="MGI:2401399"
/db_xref="taxon:10090"
/clone="9930028B14"
/sex="female"
/tissue_type="vagina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

misc_feature 29. .1968
/note="dystrophin, muscular dystrophy (MGD|MGI:94909, GB|NM_007868, evidence: BLASTN, 100%, match=999) putative"

polyA_signal 4417. .4422
/note="putative"

polyA_site 4437
/note="putative"

ORIGIN

Query Match 37.7%; Score 375.8; DB 11; Length 4437;
Best Local Similarity 96.2%; Pred. No. 1e-57;
Matches 407; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       1869 AGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 1928

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1929 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 1988

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1989 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 2048

Qy        630 AAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              |||| |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       2049 AAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTC 2108

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 748
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2109 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 2168

Qy        749 AAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 808
              ||||||| ||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       2169 AAAGTTAAAAGAAA---ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 2225

Qy        809 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAG 868
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db       2226 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACT 2285

Qy        869 GAA 871
              |||
Db       2286 AAA 2288
```

RESULT 5
CB750007
LOCUS CB750007 439 bp mRNA linear EST 11-APR-2003
DEFINITION AMGNNUC:SRPB2-00166-B9-A srpb2 (10220) Rattus norvegicus cDNA clone srpb2-00166-b9 5', mRNA sequence.
ACCESSION CB750007
VERSION CB750007.1 GI:29817309
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 439)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00166 row: b column: 9.

FEATURES Location/Qualifiers

source 1. .439
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srpb2-00166-b9"
/tissue_type="prostate tissue"
/clone_lib="srpb2 (10220)"
/note="Vector: pSPORT1; Site_1: Sal1; Site_2: NotI; rat prostate normalized double selected poly(A+) mRNA size fraction > 1 kb"

ORIGIN

Query Match 37.2%; Score 370.4; DB 14; Length 439;
Best Local Similarity 95.4%; Pred. No. 2.4e-56;
Matches 393; Conservative 0; Mismatches 16; Indels 3; Gaps 1;

```
Qy        460 GTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCA 519
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||
Db         12 GTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTTTCA 71

Qy        520 GTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGT 579
              ||||||| |||||||||||||||||||||||||||||||||||||  ||||||||||
Db         72 GTCATGATAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGACCCCCGCATGGT 131

Qy        580 TTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCT 639
              |||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db        132 TTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGCGTTTACAAGAAATAAATCT 191

Qy        640 ATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTA 699
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        192 ATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTA 251

Qy        700 TTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAG 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        252 TTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTAAAAG 311

Qy        760 AAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCAT 819
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        312 AAA---ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCAT 368

Qy        820 TTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              ||||||||||||||||||||||||||||||||||||||||||| |   ||  ||
Db        369 TTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 420


RESULT 6
BQ927942
LOCUS       BQ927942                924 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT_8858741 NCI_CGAP_Mam2 Mus musculus cDNA clone
            IMAGE:6441872 5', mRNA sequence.
ACCESSION   BQ927942
VERSION     BQ927942.1  GI:22342973
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13965  row: m  column: 09
            High quality sequence start: 10
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6441872"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN

  Query Match             35.5%;  Score 354;  DB 13;  Length 924;
  Best Local Similarity   95.3%;  Pred. No. 1.5e-53;
  Matches  387;  Conservative    0;  Mismatches   15;  Indels    4;  Gaps    2;

Qy        467 GTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGA 526
              | | ||| | |||||||||||||||||||||||||||||||||||||| ||||||||||
Db         33 GGCCTTTTCCCATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTTTCAGTCATGA 92

Qy        527 CAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 CAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATA 152

Qy        587 ATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAAT-AAATCTATATTT 645
              |||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        153 ATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAATCTATATTT 212

Qy        646 TTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTT 705
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        213 TTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTT 272

Qy        706 TGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACT 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        273 TGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTAAAAGAAA--- 329

Qy        766 ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGG 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGG 389

Qy        826 TTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||||||||||||||||||||||||||| |   ||  ||
Db        390 TTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 435


RESULT 7
AL712236
LOCUS       AL712236                629 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686J0287_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686J0287 5', mRNA sequence.
ACCESSION   AL712236
VERSION     AL712236.2  GI:32026719
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
            Wellenreuther,R., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     On Mar 22, 2002 this sequence version replaced gi:19695591.
            Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686J0287) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686J0287"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             34.4%;  Score 342.8;  DB 9;  Length 629;
  Best Local Similarity   98.0%;  Pred. No. 1.7e-51;
  Matches  347;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        518 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 60

Qy        578 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 120

Qy        638 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 180
```

```
Qy        698 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 240

Qy        758 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 817
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 300

Qy        818 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||||||||||||||||||||||| |   ||   ||
Db        301 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 354
```

RESULT 8
BX509925
LOCUS BX509925 629 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp686J2487_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686J2487 5', mRNA sequence.
ACCESSION BX509925
VERSION BX509925.1 GI:32049928
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 629)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone (DKFZp686J2487) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686J2487"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match            34.4%;  Score 342.8;  DB 13;  Length 629;
  Best Local Similarity  98.0%;  Pred. No. 1.7e-51;
  Matches  347;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        518 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 60

Qy        578 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 120

Qy        638 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 180

Qy        698 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 240

Qy        758 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 817
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 300

Qy        818 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||||||||||||||||||||||| |   ||   ||
Db        301 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 354
```

RESULT 9
BU337294
LOCUS BU337294 633 bp mRNA linear EST 28-NOV-2002
DEFINITION 603515757F1 CSEQCHN66 Gallus gallus cDNA clone ChEST455a15 5', mRNA sequence.
ACCESSION BU337294
VERSION BU337294.1 GI:25845295
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 633)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

```
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST455a15"
                     /dev_stage="16 day embryo"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN66"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match            29.7%;  Score 296;  DB 13;  Length 633;
  Best Local Similarity  87.2%;  Pred. No. 3.8e-43;
  Matches  396;  Conservative    0;  Mismatches   45;  Indels   13;  Gaps    6;
```

```
Qy        431 TTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATT 490
              ||||||  ||||  |||||||||  ||||||||||  ||  ||||||||||||||||||| |
Db        140 TTCTTTGCCTTTCTTTTCCAGGCCACAATGTAGGGAGCCTTTTCCACATGGCAGATGACT 199

Qy        491 TGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAT--- 547
              ||||  ||||||||||||  |||||||  | ||  |||||  ||  ||   |  |     |||||
Db        200 TGGGAAGAGCGATGGAAACCTTAGTGACGGTGATGACTGACGACAAGGTGTTAGAATAGC 259

Qy        548 ---AAATGTTTTACAAC-TCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAG 603
                 |  ||||||||||| ||   |  |||||||||||||||||||||||||||||||| ||||||
Db        260 AAGATGTGTTTTACAACTTCAGGCGTCCCGCATGGTTTTTATAATATTCATACTACAAAG 319

Qy        604 AGGATTAGACAGTAAGAGTTTACAAG-AAATAAATCTATATTTTTGTGAAGGGTAGTGGT 662
              ||||||||||  ||||||||||||||| ||| ||||||||||||||||||||||||||||
Db        320 AGGATTAGACTGTAAGAGTTTACAAGAAAACAAATCTATATTTTTGTGAAGGGTAGTGGT 379

Qy        663 ATTATACTGTAGATTTCAGTAGTTT-CTAAGTCTGTTATTGTTTTGTTAACAATGGCAGG 721
              | ||||||||||||||||||||||| ||||||||||  ||||||||||||||||||||||
Db        380 ACTATACTGTAGATTTCAGTAGTTTCCTAAGTCTGTTTCTGTTTTGTTAACAATGGCAGG 439

Qy        722 TTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGA 781
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        440 TTTTACACGTCTATGCAATTGTACAAAAA--TTATAAGAAAACTACATGTAAAATCTTGA 497

Qy        782 TAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTT--AAAAATTT 839
              |||||  ||||||||||||||||||||||||||||||||||  |||||||||  ||||||||
Db        498 TAGCTGAATAACTTGCCATTTCTTTATATGGAACGCATTTCGGGTTGTTTAAAAAAATTT 557

Qy        840 ATAACAGTTATAAAGAAAGAATTATAAAGGAAAA 873
              ||||||||||||||||||||     ||   ||||
Db        558 ATAACAGTTATAAAGAAAGATCGTAAACTAAAAA 591
```

RESULT 10
BI281078/c

```
LOCUS       BI281078                531 bp    mRNA    linear   EST 19-JUL-2001
DEFINITION  UI-R-DD0-bzr-a-02-0-UI.s1 UI-R-DD0 Rattus norvegicus cDNA clone
            UI-R-DD0-bzr-a-02-0-UI 3', mRNA sequence.
ACCESSION   BI281078
VERSION     BI281078.1  GI:14930459
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized fundus library cDNA Library Preparation: M.B. Soares
            Lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-DD0-bzr-a-02-0-UI"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-DD0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DD0
                     library is a non-normalized library constructed from rat
                     fundus tissue. For a detailed description of the library
                     from which this clone was derived, please visit our web
                     site at ratest.eng.uiowa.edu. The subtraction has been
                     previously described in (Bonaldo, Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_TISSUE=fundus
                     TAG_LIB=UI-R-DD0
                     TAG_SEQ=TTCGG"
ORIGIN
```

```
  Query Match           29.2%;  Score 291.2;  DB 12;  Length 531;
  Best Local Similarity   95.8%;  Pred. No. 2.9e-42;
  Matches  299;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        312 AGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 253

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||  |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        252 CTTAGTTTCAGTCATGATAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGACC 193

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db        192 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGCGTTTACAAG 133

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        132 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCT 73

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 13

Qy        750 AAGTTATAAGAA 761
              ||   | || ||
Db         12 AAAAAAAAAAAA 1
```

RESULT 11
BU226117

```
LOCUS       BU226117                854 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  603947264F1 CSEQCHN23 Gallus gallus cDNA clone ChEST901p15 5', mRNA
            sequence.
ACCESSION   BU226117
VERSION     BU226117.1  GI:25462707
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
```

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

```
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST901p15"
                     /dev_stage="22"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN23"
                     /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN
```

```
  Query Match             28.7%;  Score 286.2;  DB 13;  Length 854;
  Best Local Similarity   86.3%;  Pred. No. 1.8e-41;
  Matches  389;  Conservative    0;  Mismatches   48;  Indels   14;  Gaps    6;

Qy        437 CTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCA 496
              |||   |||    ||| |||||||||||| || ||||||||||||||||||| ||||| |
Db          5 CTCGAGGTTCCAGAGGCCACAATGTAGGGAGCCTTTTCCACATGGCAGATGACTTGGGAA 64

Qy        497 GAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAT------AAA 550
              |||||||||| |||||||  | || ||||| || ||  | |    |||||       |
Db         65 GAGCGATGGAAACCTTAGTGACGGTGATGACTGACGACAAGGTGTTAGAATAGCAAGATG 124

Qy        551 TGTTTTACAAC-TCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATT 609
              |||||||||||  ||   | |||||||||||||||||||||||||||| ||||||||||
Db        125 TGTTTTACAACTTCAGGCGTCCCGCATGGTTTTTATAATATTCATACTACAAAGAGGATT 184

Qy        610 AGACAGTAAGAGTTTACAAG-AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATA 668
              |||| |||||||||||||||| ||| ||||||| |||||||||||||||||||||| ||||
Db        185 AGACTGTAAGAGTTTACAAGAAAACAAATCTACATTTTTGTGAAGGGTAGTGGTACTATA 244

Qy        669 CTGTAGATTTCAGTAGTTT-CTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTAC 727
              |||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db        245 CTGTAGATTTCAGTAGTTTCCTAAGTCTGTTTCTGTTTTGTTAACAATGGCAGGTTTTAC 304

Qy        728 ACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTA 787
              |||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db        305 ACGTCTATGCAATTGTACAAAAA--TTATAAGAAAACTACATGTAAAATCTTGATAGCTG 362

Qy        788 AATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTT---AAAAATTTATAAC 844
              |||||||||||||||||||||||||||||||||| |||||||||   |||||||||||||
Db        363 AATAACTTGCCATTTCTTTATATGGAACGCATTTCGGGTTGTTTAAAAAAAATTTATAAC 422

Qy        845 AGTTATAAAGAAAGAATTATAAAGGAAAAAG 875
              |||||||||||   ||| ||||  ||||||
Db        423 AGTTATAAAGAACAGATTGTAAACTAAAAAG 453
```

```
RESULT 12
BU368789
LOCUS       BU368789                739 bp    mRNA    linear   EST 28-NOV-2002
DEFINITION  603596716F1 CSEQCHN73 Gallus gallus cDNA clone ChEST565o3 5', mRNA
            sequence.
ACCESSION   BU368789
VERSION     BU368789.1  GI:25876790
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .739
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST565o3"
                     /sex="Female"
                     /tissue_type="not cerebrum or cerebellum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN73"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN
```

```
  Query Match             28.4%;  Score 282.8;  DB 13;  Length 739;
  Best Local Similarity   86.5%;  Pred. No. 7.9e-41;
  Matches  384;  Conservative    0;  Mismatches   47;  Indels   13;  Gaps    6;

Qy        443 GTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGA 502
              |||   ||| |||||||||| || |||||||||||||||||| ||||| |||||||
Db        256 GTTCCAGAGGCCACAATGTAGGGAGCCTTTTCCACATGGCAGATGACTTGGGAAGAGCGA 315

Qy        503 TGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAT------AAATGTTTT 556
              ||||  ||||||  | || ||||| || ||  | |    |||||       | ||||||
Db        316 TGGAAACCTTAGTGACGGTGATGACTGACGACAAGGTGTTAGAATAGCAAGATGTGTTTT 375

Qy        557 ACAAC-TCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAG 615
              ||||| ||  |  |||||||||||||||||||||||||||||||| ||||||||||| |
Db        376 ACAACTTCAGGCGTCCCGCATGGTTTTTATAATATTCATACTACAAAGAGGATTAGACTG 435

Qy        616 TAAGAGTTTACAAG-AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAG 674
              |||||||||||||| ||| |||||||||||||||||||||||||||||| |||||||||
Db        436 TAAGAGTTTACAAGAAAACAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAG 495

Qy        675 ATTTCAGTAGTTT-CTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCT 733
              |||||||||||||| |||||||||||  ||||| ||||||||||||||||||||||||
Db        496 ATTTCAGTAGTTTCCTAAGTCTGTTTCTGTTTGGTTAACAATGGCAGGTTTTACACGTCT 555
```

```
Qy        734 ATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAAC 793
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||| ||||||
Db        556 ATGCAATTGTACAAAAA--TTATAAGAAAACTACATGTAAAATCTTGATAGCTGAATAAC 613

Qy        794 TTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTA--AAAATTTATAACAGTTATA 851
              |||||||||||||||||||||||||||||||||||||| |||||||||  ||||||||||||||||||
Db        614 TTGCCATTTCTTTATATGGAACGCATTTCGGGTTGTTTAACAAAATTTATAACAGTTATA 673

Qy        852 AAGAAAGAATTATAAAGGAAAAAG 875
              ||||||||     ||   ||||||
Db        674 AAGAAAGATCGTAAACTAAAAAAG 697


RESULT 13
AI008807/c
LOCUS       AI008807                500 bp    mRNA    linear   EST 25-JAN-1999
DEFINITION  EST203258 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMBD56 3' end, mRNA sequence.
ACCESSION   AI008807
VERSION     AI008807.1  GI:3222639
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Rattus sp."
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):2016992"
                     /db_xref="taxon:10118"
                     /clone="REMBD56"
                     /dev_stage="embryo 8, 12, 18 dpc"
                     /clone_lib="Normalized rat embryo, Bento Soares"
                     /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
ORIGIN

  Query Match             28.4%;  Score 282.6;  DB 9;  Length 500;
  Best Local Similarity   95.4%;  Pred. No. 1e-40;
  Matches  291;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy        439 CTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 498
              | | |||    |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        305 CCTAGTTCAAGAGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGA 246

Qy        499 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTAC 558
              ||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db        245 GCGATGGAGTCCTTAGTTTCAGTCATGATAGATGAAGAAGGAGCAGAATAAATGTTTTAC 186

Qy        559 AACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAA 618
              |||||||||  ||||||||||||||||||||||| ||||||||||||||||||||||||
Db        185 AACTCCTGACCCCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAA 126

Qy        619 GAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTT 678
              | |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        125 GCGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTT 66

Qy        679 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCA 6

Qy        739 ATTGT 743
              |||||
Db          5 ATTGT 1


RESULT 14
BG795230/c
LOCUS       BG795230                576 bp    mRNA    linear   EST 16-MAY-2001
DEFINITION  UTSW_SM30E11 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
            cDNA clone UTSW_SM30E11, mRNA sequence.
ACCESSION   BG795230
VERSION     BG795230.1  GI:14130800
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Gallardo,T.D., Schageman,J.J., Pertsemlidis,A., Garner,H.R.,
            Williams,R.S. and Shohet,R.V.
  TITLE     UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA
            Library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Schageman JJ
            Shohet/Garner Labs
            University of Texas Southwestern Medical Center
            6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
            Tel: 214 648 1674
            Email: Jeff.Schageman@UTSouthwestern.edu
            cDNA library constructed by UTSW as a component of the Program for
            Genomic Applications (PGA) and the Reynolds Heart Disease
            Prevention grants for use in cDNA microarray experiments. Sequence
            Quality: Sequence ends were trimmed based on percentage of ambigu
            us base calls or 'N's in windowed segments. Sequencing: First-pass
            sequencing; ABI Prism 377 sequencer and analysis software.
            Seq primer: M13/pUC Reverse.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="UTSW_SM30E11"
                     /sex="Pooled"
                     /tissue_type="Diaphragm/Hind limb muscles"
                     /cell_type="Skeletal muscle"
                     /dev_stage="2 months"
                     /lab_host="DH5a"
                     /clone_lib="UTSW Adult Mouse Skeletal Muscle Library"
                     /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
                     Primer: Oligo dT. RNA Isolation: cytoplasmic  RNA preps
                     (Manniatis); Cloning Technique: CUA Cloning (CloneAmp,
                     Life  Technologies); Average insert size: 1.8 Kb;
                     Insertion site:  TACGTCCACTGAATTCTGAGTG--->.  Other
                     information regarding entire library  may be found at
                     http://pga.swmed.edu/Data/Libraries/microarray_cdna_librar
                     ies.htm."
ORIGIN

  Query Match             28.2%;  Score 281.2;  DB 12;  Length 576;
  Best Local Similarity   95.8%;  Pred. No. 1.7e-40;
  Matches  297;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||| | |||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db        347 AGGNCNCAATGTAGGAAGCCTTTTCCACATGGCANATGATTTGGGCAGAGCGATGGAGTC 288

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 228
```

```
Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||| ||||
Db        227 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTNCAAG 168

Qy        630 AAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        167 AAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTC 108

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 748
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 48

Qy        749 AAAGTTATAA 758
              |||   | ||
Db         47 AAAAAAAAAA 38
```

RESULT 15
BY635064
LOCUS BY635064 421 bp mRNA linear EST 15-DEC-2002
DEFINITION BY635064 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K430066I18 3', mRNA sequence.
ACCESSION BY635064
VERSION BY635064.1 GI:26970246
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 421)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .421
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430066I18"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN

Query Match 28.0%; Score 278.6; DB 13; Length 421;
Best Local Similarity 98.3%; Pred. No. 5.6e-40;
Matches 292; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        125 AGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 184

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 244

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||| ||||
Db        245 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 304

Qy        630 AAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              |||| |||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        305 AAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTC 364

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTAC 745
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTAC 421
```

Search completed: April 6, 2004, 14:12:07
Job time : 3679.56 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:33:41 ; Search time 2486.92 Seconds
(without alignments)
1500.955 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2470632 seqs, 1873875610 residues

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	996	100.0	996	9	US-09-966-264-2	Sequence 2, Appli
2	410.8	41.2	5339	15	US-10-149-736-40	Sequence 40, Appl
3	410.8	41.2	5417	15	US-10-149-736-39	Sequence 39, Appl
4	410.8	41.2	5462	15	US-10-149-736-41	Sequence 41, Appl
5	410.8	41.2	8689	15	US-10-149-736-42	Sequence 42, Appl
6	410.8	41.2	11443	15	US-10-149-736-44	Sequence 44, Appl
7	410.8	41.2	12057	15	US-10-149-736-47	Sequence 47, Appl
8	410.8	41.2	13957	9	US-09-782-378A-22	Sequence 22, Appl
9	410.8	41.2	13957	9	US-09-880-107-2284	Sequence 2284, Ap
10	410.8	41.2	13957	15	US-10-149-736-1	Sequence 1, Appli
11	410.8	41.2	14069	12	US-10-342-887-434	Sequence 434, App
12	410.8	41.2	14082	12	US-10-342-887-981	Sequence 981, App
13	410.8	41.2	14082	15	US-10-341-434-108	Sequence 108, App
14	396.8	39.8	2691	15	US-10-149-736-38	Sequence 38, Appl
15	374.2	37.6	13815	15	US-10-149-736-2	Sequence 2, Appli
c 16	151.4	15.2	430	9	US-09-796-692-3505	Sequence 3505, Ap
c 17	151.4	15.2	430	14	US-10-040-862-3505	Sequence 3505, Ap
c 18	151.4	15.2	430	15	US-10-057-475B-3505	Sequence 3505, Ap
c 19	151.4	15.2	430	15	US-10-154-884B-3505	Sequence 3505, Ap
c 20	151.4	15.2	2691	15	US-10-149-736-38	Sequence 38, Appl
c 21	151.4	15.2	8689	15	US-10-149-736-42	Sequence 42, Appl
c 22	151.4	15.2	11443	15	US-10-149-736-44	Sequence 44, Appl
c 23	151.4	15.2	13957	9	US-09-782-378A-22	Sequence 22, Appl
c 24	151.4	15.2	13957	9	US-09-880-107-2284	Sequence 2284, Ap
c 25	151.4	15.2	13957	15	US-10-149-736-1	Sequence 1, Appli
c 26	151.4	15.2	14069	12	US-10-342-887-434	Sequence 434, App
c 27	151.4	15.2	14082	12	US-10-342-887-981	Sequence 981, App
c 28	151.4	15.2	14082	15	US-10-341-434-108	Sequence 108, App
29	137	13.8	137	9	US-09-966-264-1	Sequence 1, Appli
c 30	101.4	10.2	13815	15	US-10-149-736-2	Sequence 2, Appli
c 31	51.2	5.1	3673778	14	US-10-312-841-1	Sequence 1, Appli
32	50.6	5.1	12393	14	US-10-311-455-1236	Sequence 1236, Ap
33	49.8	5.0	18154	14	US-10-311-455-228	Sequence 228, App
c 34	49.4	5.0	622	12	US-10-424-599-3380	Sequence 3380, Ap
35	48.4	4.9	23683	14	US-10-240-485-176	Sequence 176, App
c 36	48.2	4.8	5546	14	US-10-311-455-364	Sequence 364, App
37	48	4.8	8076	16	US-10-257-166-36	Sequence 36, Appl
c 38	47.4	4.8	5413	12	US-10-221-714A-418	Sequence 418, App
39	47.4	4.8	3673778	14	US-10-312-841-1	Sequence 1, Appli
40	47.4	4.8	3673778	14	US-10-312-841-2	Sequence 2, Appli
41	47.2	4.7	17213	14	US-10-311-455-1455	Sequence 1455, Ap
42	47	4.7	4993	14	US-10-311-455-2075	Sequence 2075, Ap
43	47	4.7	7348	12	US-10-221-714A-57	Sequence 57, Appl
44	47	4.7	8771	14	US-10-311-455-1797	Sequence 1797, Ap
45	46.8	4.7	5683	14	US-10-311-455-1780	Sequence 1780, Ap

ALIGNMENTS

RESULT 1
US-09-966-264-2
; Sequence 2, Application US/09966264
; Patent No. US20020099015A1
; GENERAL INFORMATION:
; APPLICANT: Barber, Elizabeth K
; TITLE OF INVENTION: Gene Expression Control Element DNA
; FILE REFERENCE: 896034605001
; CURRENT APPLICATION NUMBER: US/09/966,264
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/237,079
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 996
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(996)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (710)..(996)
; OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph
; OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a
; OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence
; OTHER INFORMATION: in the dystrophin DNA
US-09-966-264-2

Query Match 100.0%; Score 996; DB 9; Length 996;
Best Local Similarity 100.0%; Pred. No. 6.9e-224;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
```

```
Qy        61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120

Qy       121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180

Qy       181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240

Qy       241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300

Qy       301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360

Qy       361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420

Qy       421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480

Qy       481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540

Qy       541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600

Qy       601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660

Qy       661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720

Qy       721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780

Qy       781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840

Qy       841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900

Qy       901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960

Qy       961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
             ||||||||||||||||||||||||||||||||||||
Db       961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996


RESULT 2
US-10-149-736-40
; Sequence 40, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
;   LENGTH: 5339
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-40

  Query Match             41.2%;  Score 410.8;  DB 15;  Length 5339;
  Best Local Similarity   98.3%;  Pred. No. 5.7e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4551 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4610

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4611 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4730

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4731 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4790

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4791 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4850

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| |  ||
Db      4851 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 4910

Qy       870 AA 871
             ||
Db      4911 AA 4912


RESULT 3
US-10-149-736-39
; Sequence 39, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
;   LENGTH: 5417
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-39

  Query Match             41.2%;  Score 410.8;  DB 15;  Length 5417;
  Best Local Similarity   98.3%;  Pred. No. 5.7e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4628

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4629 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4688

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4689 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4748

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4749 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4808

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4868

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4869 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4928

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 4988

Qy        870 AA 871
              ||
Db       4989 AA 4990


RESULT 4
US-10-149-736-41
; Sequence 41, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
;   LENGTH: 5462
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-41

  Query Match             41.2%;  Score 410.8;  DB 15;  Length 5462;
  Best Local Similarity   98.3%;  Pred. No. 5.8e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4673

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4733

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4734 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4793

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4794 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4853

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4854 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4913

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4914 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4973

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4974 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 5033

Qy        870 AA 871
              ||
Db       5034 AA 5035


RESULT 5
US-10-149-736-42
; Sequence 42, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 8689
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-42

  Query Match             41.2%;  Score 410.8;  DB 15;  Length 8689;
  Best Local Similarity   98.3%;  Pred. No. 7.3e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 6044

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 6104

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6105 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 6164
```

```
Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6165 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 6224

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6225 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 6284

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6285 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 6344

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db       6345 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 6404

Qy        870 AA 871
              ||
Db       6405 AA 6406
```

RESULT 6
US-10-149-736-44
; Sequence 44, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 11443
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-44

Query Match 41.2%; Score 410.8; DB 15; Length 11443;
Best Local Similarity 98.3%; Pred. No. 8.4e-86;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8739 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 8798

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 8858

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8859 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 8918

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8919 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 8978

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8979 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 9038

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9039 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 9098

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db       9099 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 9158

Qy        870 AA 871
              ||
Db       9159 AA 9160
```

RESULT 7
US-10-149-736-47
; Sequence 47, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 12057
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-47

Query Match 41.2%; Score 410.8; DB 15; Length 12057;
Best Local Similarity 98.3%; Pred. No. 8.6e-86;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11214 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11273

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11274 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11333

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11334 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11393

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11394 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11453

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11454 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11513

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11514 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11573

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11574 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11633

Qy        870 AA 871
              ||
Db      11634 AA 11635
```

RESULT 8

US-09-782-378A-22
; Sequence 22, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie
; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-782-378A-22

Query Match 41.2%; Score 410.8; DB 9; Length 13957;
Best Local Similarity 98.3%; Pred. No. 9.3e-86;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

RESULT 9
US-09-880-107-2284
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284

Query Match 41.2%; Score 410.8; DB 9; Length 13957;
Best Local Similarity 98.3%; Pred. No. 9.3e-86;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

RESULT 10
US-10-149-736-1
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-1

```
  Query Match             41.2%;  Score 410.8;  DB 15;  Length 13957;
  Best Local Similarity   98.3%;  Pred. No. 9.3e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |  ||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11672

Qy        870 AA 871
              ||
Db      11673 AA 11674
```

RESULT 11
US-10-342-887-434
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
; LENGTH: 14069
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-434

```
  Query Match             41.2%;  Score 410.8;  DB 12;  Length 14069;
  Best Local Similarity   98.3%;  Pred. No. 9.3e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11365 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11424

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11425 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11484

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11485 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11544

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11545 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11604

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11605 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11664

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11665 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11724

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |  ||
Db      11725 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11784

Qy        870 AA 871
              ||
Db      11785 AA 11786
```

RESULT 12
US-10-342-887-981
; Sequence 981, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 981
; LENGTH: 14082
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-981

```
  Query Match             41.2%;  Score 410.8;  DB 12;  Length 14082;
  Best Local Similarity   98.3%;  Pred. No. 9.4e-86;
  Matches  415;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11378 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11437

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11438 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11497
```

```
Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11498 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11557

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11558 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11617

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11618 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11677

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11678 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11737

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              |||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11738 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11797

Qy        870 AA 871
              ||
Db      11798 AA 11799
```

RESULT 13
US-10-341-434-108
; Sequence 108, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 108
; LENGTH: 14082
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (703)..(11388)
; OTHER INFORMATION:
US-10-341-434-108

Query Match 41.2%; Score 410.8; DB 15; Length 14082;
Best Local Similarity 98.3%; Pred. No. 9.4e-86;
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11378 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11437

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11438 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11497

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11498 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11557

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11558 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11617

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11618 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11677

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11678 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11737

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGG 869
              |||||||||||||||||||||||||||||||||||||||||||||||||| |   ||
Db      11738 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTA 11797

Qy        870 AA 871
              ||
Db      11798 AA 11799
```

RESULT 14
US-10-149-736-38
; Sequence 38, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 2691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-38

Query Match 39.8%; Score 396.8; DB 15; Length 2691;
Best Local Similarity 98.3%; Pred. No. 7.8e-83;
Matches 401; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy        464 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 60

Qy        524 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 120

Qy        584 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 180

Qy        644 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 240

Qy        704 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 300

Qy        764 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 360

Qy        824 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAGGAA 871
              |||||||||||||||||||||||||||||||||||| |   ||   ||
Db        361 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAA 408
```

RESULT 15
US-10-149-736-2

```
; Sequence 2, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 13815
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-149-736-2

  Query Match           37.6%;  Score 374.2;  DB 15;  Length 13815;
  Best Local Similarity   96.0%;  Pred. No. 3.7e-77;
  Matches  406;  Conservative    0;  Mismatches   13;  Indels    4;  Gaps    2;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      11244 AGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11303

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11304 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11363

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      11364 CCCGCATGGTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11423

Qy        630 AAAT-AAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              |||| |||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      11424 AAATAAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTC 11483

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 748
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11484 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 11543

Qy        749 AAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 808
              ||||||| ||||||   |||||||||||||||||||||||||||||||||||||||||||
Db      11544 AAAGTTAAAAGAAA---ACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 11600

Qy        809 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGAATTATAAAG 868
              |||||||||||||||||||||||||||||||||||||||||||| ||| |   ||
Db      11601 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAGAGATTGTAAACT 11660

Qy        869 GAA 871
              ||
Db      11661 AAA 11663


Search completed: April  6, 2004, 15:01:46
Job time : 2493.92 secs
```

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:12:12 ; Search time 3672.8 Seconds
(without alignments)
8098.097 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	410	41.2	844	12	BG706268	BG706268 602669567
	2	342	34.3	629	9	AL712236	AL712236 DKFZp686J
	3	342	34.3	629	13	BX509925	BX509925 DKFZp686J
	4	238	23.9	999	13	BX419180	BX419180 BX419180
	5	231	23.2	739	9	AU138648	AU138648 AU138648
c	6	221	22.2	308	14	N75050	N75050 za89c07.s1
	7	209	21.0	275	10	AW604756	AW604756 CM0-CT034
	8	209	21.0	277	10	AW861957	AW861957 CM3-CT034
	9	209	21.0	284	10	AW858457	AW858457 CM3-CT034
	10	208	20.9	888	14	CF552038	CF552038 AGENCOURT
	11	202	20.3	525	28	AQ679243	AQ679243 HS_5417_A
c	12	194	19.5	271	10	AW858578	AW858578 CM3-CT034
c	13	183	18.4	619	10	AW771158	AW771158 hn53c06.x
	14	161	16.2	309	14	H27701	H27701 yl61c12.r1
c	15	148	14.9	556	28	AQ504173	AQ504173 RPCI-11-2
c	16	146	14.7	396	14	H89576	H89576 yw28c12.r1
c	17	146	14.7	455	9	AA427831	AA427831 zw49d12.r
c	18	146	14.7	462	14	CB045405	CB045405 NISC_gc10
c	19	146	14.7	473	14	CA389721	CA389721 cs101f03.
c	20	146	14.7	489	28	AQ015185	AQ015185 CIT-HSP-2
c	21	146	14.7	647	10	AW664684	AW664684 hi84e10.x
c	22	146	14.7	696	10	AW950480	AW950480 EST362550
c	23	146	14.7	710	9	AV725574	AV725574 AV725574
c	24	146	14.7	727	12	BG567176	BG567176 602589640
c	25	146	14.7	740	14	CD356811	CD356811 AGENCOURT
c	26	146	14.7	745	14	CD357556	CD357556 AGENCOURT
c	27	146	14.7	1121	12	BM546012	BM546012 AGENCOURT
c	28	145	14.6	743	14	CD110642	CD110642 AGENCOURT
c	29	141	14.2	373	10	AW580404	AW580404 PM2-HT045
	30	129	13.0	939	13	BQ718541	BQ718541 AGENCOURT
	31	120	12.0	352	10	AW385154	AW385154 PM2-HT045
c	32	118	11.8	501	13	BU619605	BU619605 UI-H-FH1-
c	33	112	11.2	862	14	CB962272	CB962272 AGENCOURT
	34	111	11.1	353	10	AW580423	AW580423 PM2-HT045
	35	109	10.9	183	10	AW858537	AW858537 CM3-CT034
c	36	97	9.7	911	12	BI752714	BI752714 603028327
	37	94	9.4	439	14	CB750007	CB750007 AMGNNUC:S
	38	94	9.4	924	13	BQ927942	BQ927942 AGENCOURT
	39	94	9.4	4437	11	AK036936	AK036936 Mus muscu
	40	91	9.1	492	14	CB725360	CB725360 AMGNNUC:N
	41	91	9.1	571	14	CA893354	CA893354 B0178E05-
c	42	90	9.0	375	10	AW607064	AW607064 PM2-HT045
	43	88	8.8	484	10	BF655326	BF655326 279797 MA
c	44	87	8.7	531	12	BI281078	BI281078 UI-R-DD0-
c	45	87	8.7	576	12	BG795230	BG795230 UTSW_SM30

ALIGNMENTS

RESULT 1
BG706268
LOCUS BG706268 844 bp mRNA linear EST 07-MAY-2001
DEFINITION 602669567F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4792313 5', mRNA sequence.
ACCESSION BG706268
VERSION BG706268.1 GI:13981445
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10670 row: e column: 18

```
          High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4792313"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             41.2%;  Score 410;  DB 12;  Length 844;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 147

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 207

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 267

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 327

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 387

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 447

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        448 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 497


RESULT 2
AL712236
LOCUS       AL712236                629 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686J0287_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686J0287 5', mRNA sequence.
ACCESSION   AL712236
VERSION     AL712236.2  GI:32026719
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
            Wellenreuther,R., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     On Mar 22, 2002 this sequence version replaced gi:19695591.
            Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686J0287) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686J0287"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             34.3%;  Score 342;  DB 9;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 3e-156;
  Matches  342;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        518 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 60

Qy        578 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 120

Qy        638 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 180

Qy        698 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 240

Qy        758 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 817
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 300

Qy        818 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 342


RESULT 3
BX509925
LOCUS       BX509925                629 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686J2487_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686J2487 5', mRNA sequence.
ACCESSION   BX509925
VERSION     BX509925.1  GI:32049928
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
```

```
  TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
            Wellenreuther,R., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686J2487) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686J2487"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             34.3%;  Score 342;  DB 13;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 3e-156;
  Matches  342;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        518 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATG 60

Qy        578 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 120

Qy        638 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGT 180

Qy        698 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATA 240

Qy        758 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 817
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGC 300

Qy        818 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 342


RESULT 4
BX419180
LOCUS       BX419180                999 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX419180 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF014YA17 5-PRIME, mRNA sequence.
ACCESSION   BX419180
VERSION     BX419180.1  GI:30637926
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 999)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10533.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF014AA09QP1&cluster=10533.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF014AA09QP1.
FEATURES             Location/Qualifiers
     source          1. .999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF014YA17"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             23.9%;  Score 238;  DB 13;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-105;
  Matches  238;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        622 TTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAG 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 TTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAG 128

Qy        682 TAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATT 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 TAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATT 188

Qy        742 GTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATT 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATT 248

Qy        802 TCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 TCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 306


RESULT 5
AU138648
LOCUS       AU138648                739 bp    mRNA    linear   EST 02-AUG-2002
DEFINITION  AU138648 PLACE1 Homo sapiens cDNA clone PLACE1009015 5', mRNA
            sequence.
ACCESSION   AU138648
VERSION     AU138648.1  GI:11000169
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
```

```
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1. .739
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="PLACE1009015"
                      /tissue_type="placenta"
                      /clone_lib="PLACE1"
                      /note="Vector: pME18SFL3"
ORIGIN

  Query Match             23.2%;  Score 231;  DB 9;  Length 739;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-102;
  Matches  231;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        629 GAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 688
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         38 GAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTC 97

Qy        689 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 748
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         98 TAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAA 157

Qy        749 AAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 808
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 AAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTAT 217

Qy        809 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 ATGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 268


RESULT 6
N75050/c
LOCUS       N75050                 308 bp    mRNA    linear   EST 29-MAR-1996
DEFINITION  za89c07.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:299724 3' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA
            sequence.
ACCESSION   N75050
VERSION     N75050.1  GI:1237628
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: m13 -40 forward
            High quality sequence stop: 227.
FEATURES              Location/Qualifiers
     source           1. .308
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:1244648"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:299724"
                      /dev_stage="19 weeks"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares_fetal_lung_NbHL19W"
                      /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. This library was constructed
                      from the same fetus as the fetal heart library, Soares
                      fetal heart NbHH19W."
ORIGIN

  Query Match             22.2%;  Score 221;  DB 14;  Length 308;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-97;
  Matches  221;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        639 TATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTT 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 TATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTT 171

Qy        699 ATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAA 758
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 ATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAA 111

Qy        759 GAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCA 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 GAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCA 51

Qy        819 TTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||
Db         50 TTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 10


RESULT 7
AW604756
LOCUS       AW604756               275 bp    mRNA    linear   EST 23-MAR-2000
DEFINITION  CM0-CT0341-260100-160-a10 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW604756
VERSION     AW604756.1  GI:7309497
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 275)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-CT0341-
            260100-160-a10&t3=2000-01-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 275.
FEATURES              Location/Qualifiers
     source           1. .275
```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0341"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 21.0%; Score 209; DB 10; Length 275;
Best Local Similarity 100.0%; Pred. No. 4.8e-91;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         55 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 114

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 174

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 234

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAG 658
              |||||||||||||||||||||||||||||
Db        235 AAATAAATCTATATTTTTGTGAAGGGTAG 263
```

RESULT 8
AW861957
LOCUS AW861957 277 bp mRNA linear EST 19-MAY-2000
DEFINITION CM3-CT0341-190400-152-c06 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW861957
VERSION AW861957.1 GI:7957650
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 277)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-CT0341-190400-152-c06&t3=2000-04-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 277.

FEATURES Location/Qualifiers
source 1. .277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0341"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 21.0%; Score 209; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 4.8e-91;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         57 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 116

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 176

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 236

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAG 658
              |||||||||||||||||||||||||||||
Db        237 AAATAAATCTATATTTTTGTGAAGGGTAG 265
```

RESULT 9
AW858457
LOCUS AW858457 284 bp mRNA linear EST 19-MAY-2000
DEFINITION CM3-CT0341-170200-093-b02 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW858457
VERSION AW858457.1 GI:7954150
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 284)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-CT0341-170200-093-b02&t3=2000-02-17&t4=1)
Seq primer: puc 18 forward

```
          High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="CT0341"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No.
                     196,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
ORIGIN

  Query Match             21.0%;  Score 209;  DB 10;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-91;
  Matches  209;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 123

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 183

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 243

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAG 658
              |||||||||||||||||||||||||||||
Db        244 AAATAAATCTATATTTTTGTGAAGGGTAG 272


RESULT 10
CF552038
LOCUS       CF552038                888 bp    mRNA    linear   EST 22-SEP-2003
DEFINITION  AGENCOURT_15595498 NIH_MGC_183 Homo sapiens cDNA clone
            IMAGE:30529933 5', mRNA sequence.
ACCESSION   CF552038
VERSION     CF552038.1  GI:34888872
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM618  row: e  column: 14
            High quality sequence start: 17
            High quality sequence stop: 708.
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30529933"
                     /lab_host="DH10B-TonA (T1 and T5 phage resistant)"
                     /clone_lib="NIH_MGC_183"
                     /note="Organ: Pooled muscle (cardiac and skeletal);
                     Vector: pCMV-SPORT6.1; Site_1: EcoRV (destroyed); Site_2:
                     NotI; Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 1.7. Library was constructed by Invitrogen."
ORIGIN

  Query Match             20.9%;  Score 208;  DB 14;  Length 888;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-90;
  Matches  208;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        631 AATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTA 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         35 AATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTA 94

Qy        691 AGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAA 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAA 154

Qy        751 AGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATAT 810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 AGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATAT 214

Qy        811 GGAACGCATTTTGGGTTGTTTAAAAATT 838
              ||||||||||||||||||||||||||||
Db        215 GGAACGCATTTTGGGTTGTTTAAAAATT 242


RESULT 11
AQ679243
LOCUS       AQ679243                525 bp    DNA     linear   GSS 25-JUN-1999
DEFINITION  HS_5417_A1_B06_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=993 Col=11 Row=C, genomic survey sequence.
ACCESSION   AQ679243
VERSION     AQ679243.1  GI:5228047
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 993  row: C  column: 11
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Homo sapiens"
```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=993 Col=11 Row=C"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN

Query Match 20.3%; Score 202; DB 28; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.2e-87;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      217 TGATCCAACATTAAGCTGATACTAACAAACAACGTGTAATGGCTTCATTAATAAGGCTTT 276
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      103 TGATCCAACATTAAGCTGATACTAACAAACAACGTGTAATGGCTTCATTAATAAGGCTTT 162

Qy      277 GCTTCTTCCTGGAAACTGGTGAAAAATCAAACCTTGTTGTGTACACCCTCGATGCAGCTT 336
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      163 GCTTCTTCCTGGAAACTGGTGAAAAATCAAACCTTGTTGTGTACACCCTCGATGCAGCTT 222

Qy      337 CTGTGTTGTCTTCACCCAGAAATGGGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGA 396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      223 CTGTGTTGTCTTCACCCAGAAATGGGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGA 282

Qy      397 TGCTATCTATCTGCACCTTTTG 418
            |||||||||||||||||||||
Db      283 TGCTATCTATCTGCACCTTTTG 304
```

RESULT 12
AW858578/c
LOCUS AW858578 271 bp mRNA linear EST 19-MAY-2000
DEFINITION CM3-CT0341-190400-152-a10 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW858578
VERSION AW858578.1 GI:7954271
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-CT0341-190
400-152-a10&t3=2000-04-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 32
High quality sequence stop: 271.
FEATURES Location/Qualifiers
source 1. .271

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0341"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

ORIGIN

Query Match 19.5%; Score 194; DB 10; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.1e-83;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      206 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 147

Qy      510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      146 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 87

Qy      570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       86 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 27

Qy      630 AAATAAATCTATAT 643
            ||||||||||||||
Db       26 AAATAAATCTATAT 13
```

RESULT 13
AW771158/c
LOCUS AW771158 619 bp mRNA linear EST 04-MAY-2000
DEFINITION hn53c06.x1 NCI_CGAP_Co17 Homo sapiens cDNA clone IMAGE:3027370 3'
similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.
ACCESSION AW771158
VERSION AW771158.1 GI:7703211
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 619)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D.,Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 415.
FEATURES Location/Qualifiers
source 1. .619
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3027370"
/tissue_type="juvenile granulosa tumor"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Co17"

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

ORIGIN

Query Match 18.4%; Score 183; DB 10; Length 619;
Best Local Similarity 100.0%; Pred. No. 2.3e-78;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      470 TTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAG 529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      196 TTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAG 137

Qy      530 ATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATA 589
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      136 ATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATA 77

Qy      590 TTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGT 649
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       76 TTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGT 17

Qy      650 GAA 652
            |||
Db       16 GAA 14
```

RESULT 14
H27701
LOCUS H27701 309 bp mRNA linear EST 13-JUL-1995
DEFINITION yl61c12.r1 Soares breast 3NbHBst Homo sapiens cDNA clone IMAGE:162742 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.
ACCESSION H27701
VERSION H27701.1 GI:898054
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 309)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 439
High quality sequence stops: 291
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 439 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 291.

FEATURES Location/Qualifiers
source 1. .309
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:577035"
/db_xref="taxon:9606"
/clone="IMAGE:162742"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares breast 3NbHBst"
/note="Organ: breast; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 16.2%; Score 161; DB 14; Length 309;
Best Local Similarity 100.0%; Pred. No. 1.5e-67;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      699 ATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAA 758
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAA 60

Qy      759 GAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCA 818
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCA 120

Qy      819 TTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            |||||||||||||||||||||||||||||||||||||||||
Db      121 TTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 161
```

RESULT 15
AQ504173/c
LOCUS AQ504173 556 bp DNA linear GSS 29-APR-1999
DEFINITION RPCI-11-275F14.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-275F14, genomic survey sequence.
ACCESSION AQ504173
VERSION AQ504173.1 GI:4708920
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 556)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .556
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7605349"
/db_xref="taxon:9606"
/clone="RPCI-11-275F14"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 14.9%; Score 148; DB 28; Length 556;
Best Local Similarity 100.0%; Pred. No. 3.2e-61;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 94

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 34

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTC 148
              ||||||||||||||||||||||||||||
Db         33 AGAAATCTTGTGCTTATCTATGGAATTC 6
```

Search completed: April 6, 2004, 17:40:09
Job time : 3674.8 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 13:02:17 ; Search time 4176.52 Seconds
(without alignments)
10336.278 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	859	86.2	98056	9	AC006061	AC006061 Homo sapi
	2	808	81.1	212120	2	AC023414	AC023414 Homo sapi
	3	410	41.2	2463	6	AX817316	AX817316 Sequence
	4	410	41.2	4658	9	BC028720	BC028720 Homo sapi
	5	410	41.2	5339	6	AX538620	AX538620 Sequence
	6	410	41.2	5417	6	AX538619	AX538619 Sequence
	7	410	41.2	5462	6	AX538621	AX538621 Sequence
	8	410	41.2	8689	6	AX538622	AX538622 Sequence
	9	410	41.2	11443	6	AX538624	AX538624 Sequence
	10	410	41.2	12057	6	AX538627	AX538627 Sequence
	11	410	41.2	12446	9	HSDMDR	X14298 Human mRNA
	12	410	41.2	13957	6	AX409637	AX409637 Sequence
	13	410	41.2	13957	6	AX538581	AX538581 Sequence
	14	410	41.2	13957	9	HUMDYS	M18533 Homo sapien
	15	396	39.8	2148	11	G15848	G15848 human STS C
	16	396	39.8	2691	6	AX538618	AX538618 Sequence
	17	359	36.0	13977	6	AR220819	AR220819 Sequence
	18	350	35.1	350	9	AF213444	AF213444 Homo sapi
	19	289	29.0	2110	9	HUMDMDXX	M92650 Human Duche
	20	278	27.9	2563	9	AK129855	AK129855 Homo sapi
	21	260	26.1	3163	6	E30223	E30223 Shortened d
	22	260	26.1	3172	6	E30222	E30222 Shortened d
	23	260	26.1	4075	6	E30221	E30221 Shortened d
	24	260	26.1	4402	6	E30219	E30219 Shortened d
	25	260	26.1	4402	6	E30220	E30220 Shortened d
	26	251	25.2	3747	6	E30218	E30218 Shortened d
c	27	231	23.2	212120	2	AC023414	AC023414 Homo sapi
	28	218	21.9	218	9	HUMDYST20	M86903 H.sapiens d
c	29	146	14.7	2148	11	G15848	G15848 human STS C
c	30	146	14.7	2563	9	AK129855	AK129855 Homo sapi
c	31	146	14.7	2691	6	AX538618	AX538618 Sequence
c	32	146	14.7	4658	9	BC028720	BC028720 Homo sapi
c	33	146	14.7	8689	6	AX538622	AX538622 Sequence
c	34	146	14.7	11443	6	AX538624	AX538624 Sequence
c	35	146	14.7	13957	6	AX409637	AX409637 Sequence
c	36	146	14.7	13957	6	AX538581	AX538581 Sequence
c	37	146	14.7	13957	9	HUMDYS	M18533 Homo sapien
	38	146	14.7	98056	9	AC006061	AC006061 Homo sapi
	39	145	14.6	2005	6	AX817318	AX817318 Sequence
c	40	126	12.7	13977	6	AR220819	AR220819 Sequence
	41	102	10.2	13887	4	AF070485	AF070485 Canis fam
	42	94	9.4	189131	10	AL645477	AL645477 Mouse DNA
	43	94	9.4	272578	2	AC108338	AC108338 Rattus no
c	44	94	9.4	279539	2	AC114184	AC114184 Rattus no
	45	91	9.1	13815	6	AX306153	AX306153 Sequence

ALIGNMENTS

RESULT 1
AC006061/c
LOCUS AC006061 98056 bp DNA linear PRI 01-MAY-2002
DEFINITION Homo sapiens X BAC GSHB-19O24 (Genome Systems Human BAC Library) complete sequence.
ACCESSION AC006061
VERSION AC006061.1 GI:4204246
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 98056)
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,

Gorrell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L., Kondejewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 7 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 8 (bases 1 to 98056)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 30, 1999 this sequence version replaced gi:4176317.
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

FEATURES Location/Qualifiers
source 1. .98056
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="GSHB-19O24"
misc_feature 1. .1999
/note="overlaps bases 151178. .153177 of clone AC117405"
/function="clone overlap"
repeat_region 40. .82
/rpt_family="L2"
repeat_region 378. .507
/rpt_family="L1MD3"
repeat_region complement(518. .570)
/rpt_family="(GGA)n"
repeat_region 571. .729
/rpt_family="L1MC4"
repeat_region 730. .987
/rpt_family="MSTD"
repeat_region complement(991. .1400)
/rpt_family="L2"
repeat_region complement(1649. .1779)
/rpt_family="MIR"
repeat_region complement(1783. .2212)
/rpt_family="LTR16A"
repeat_region complement(3626. .3656)
/rpt_family="(GA)n"
repeat_region complement(4713. .5074)
/rpt_family="THE1B"
repeat_region 5702. .7108
/rpt_family="L2"
repeat_region complement(8670. .8734)
/rpt_family="MLT1G"
repeat_region complement(9069. .9144)
/rpt_family="MLT1G"
repeat_region complement(9254. .9351)
/rpt_family="MIR"
repeat_region complement(9537. .9688)
/rpt_family="MER5A"
repeat_region 9797. .9942
/rpt_family="MIR"
repeat_region complement(10403. .10429)
/rpt_family="AT_rich"
repeat_region complement(10673. .10705)
/rpt_family="(CA)n"
repeat_region complement(11460. .11755)
/rpt_family="AluSx"
repeat_region 11977. .12125
/rpt_family="MER5A"
repeat_region 12556. .13643
/rpt_family="L1MD2"
repeat_region 13618. .13808
/rpt_family="L1MD3"
repeat_region complement(13997. .14297)
/rpt_family="MER4D"
repeat_region 14327. .14625
/rpt_family="AluSg"
repeat_region complement(14648. .14915)
/rpt_family="MER4D"
repeat_region complement(14915. .15147)
/rpt_family="MER4D"
repeat_region 15286. .15404
/rpt_family="FLAM_C"

```
     repeat_region   16169. .16463
                     /rpt_family="AluSx"
     repeat_region   16520. .16570
                     /rpt_family="L2"
     repeat_region   16688. .16904
                     /rpt_family="L2"
     repeat_region   complement(16955. .17147)
                     /rpt_family="MIR"
     repeat_region   complement(17171. .17467)
                     /rpt_family="AluY"
     repeat_region   complement(20101. .20241)
                     /rpt_family="MIR"
     STS             22625. .22788
                     /standard_name="DXS1097"
                     /db_xref="GDB:190032"
     repeat_region   24144. .24435
                     /rpt_family="AluSq"
     repeat_region   complement(25010. .25235)
                     /rpt_family="L2"
     repeat_region   25643. .25771
                     /rpt_family="MER58A"
     repeat_region   complement(26111. .26429)
                     /rpt_family="L2"
     repeat_region   complement(26578. .26636)
                     /rpt_family="(CATA)n"
     repeat_region   complement(26956. .27027)
                     /rpt_family="L1MA9"
     repeat_region   complement(27038. .27188)
                     /rpt_family="L2"
     repeat_region   27203. .27641
                     /rpt_family="L1MA8"
     repeat_region   27642. .28179
                     /rpt_family="L1PA4"
     repeat_region   28194. .28495
                     /rpt_family="AluSx"
     repeat_region   complement(28509. .28660)
                     /rpt_family="L2"
     repeat_region   28665. .28830
                     /rpt_family="AluJb"
     repeat_region   complement(28848. .29503)
                     /rpt_family="L2"

  Query Match             86.2%;  Score 859;  DB 9;  Length 98056;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  859;  Conservative   0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94154 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 94095

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94094 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 94035

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94034 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 93975

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93974 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 93915

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93914 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 93855

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93854 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 93795

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93794 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 93735

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93734 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 93675

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93674 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 93615

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93614 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 93555

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93554 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 93495

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93494 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 93435

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93434 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 93375

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93374 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 93315

Qy        841 TAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||
Db      93314 TAACAGTTATAAAGAAAGA 93296


RESULT 2
AC023414
LOCUS       AC023414               212120 bp    DNA     linear   HTG 11-APR-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-767B5 map 14, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC023414
VERSION     AC023414.2  GI:7534009
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-767B5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212120)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 11, 2000 this sequence version replaced gi:6970579.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L6731
              Center clone name: 767_B_5
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 185633 bases at least Q40
              Consensus quality: 195115 bases at least Q30
              Consensus quality: 201837 bases at least Q20
              Insert size: 178000; agarose-fp
              Insert size: 209120; sum-of-contigs
              Quality coverage: 5.8 in Q20 bases; agarose-fp
              Quality coverage: 4.9 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 31 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1043: contig of 1043 bp in length
          *     1044     1143: gap of 100 bp
          *     1144     2224: contig of 1081 bp in length
          *     2225     2324: gap of 100 bp
          *     2325     3345: contig of 1021 bp in length
          *     3346     3445: gap of 100 bp
          *     3446     4502: contig of 1057 bp in length
          *     4503     4602: gap of 100 bp
          *     4603     5756: contig of 1154 bp in length
          *     5757     5856: gap of 100 bp
          *     5857     7097: contig of 1241 bp in length
          *     7098     7197: gap of 100 bp
          *     7198     8236: contig of 1039 bp in length
          *     8237     8336: gap of 100 bp
          *     8337     9353: contig of 1017 bp in length
          *     9354     9453: gap of 100 bp
          *     9454    10705: contig of 1252 bp in length
          *    10706    10805: gap of 100 bp
          *    10806    12332: contig of 1527 bp in length
          *    12333    12432: gap of 100 bp
          *    12433    13877: contig of 1445 bp in length
          *    13878    13977: gap of 100 bp
          *    13978    15201: contig of 1224 bp in length
          *    15202    15301: gap of 100 bp
          *    15302    16968: contig of 1667 bp in length
          *    16969    17068: gap of 100 bp
          *    17069    19060: contig of 1992 bp in length
          *    19061    19160: gap of 100 bp
          *    19161    22639: contig of 3479 bp in length
          *    22640    22739: gap of 100 bp
          *    22740    26376: contig of 3637 bp in length
          *    26377    26476: gap of 100 bp
          *    26477    29275: contig of 2799 bp in length
          *    29276    29375: gap of 100 bp
          *    29376    33265: contig of 3890 bp in length
          *    33266    33365: gap of 100 bp
          *    33366    37012: contig of 3647 bp in length
          *    37013    37112: gap of 100 bp
          *    37113    39390: contig of 2278 bp in length
          *    39391    39490: gap of 100 bp
          *    39491    43690: contig of 4200 bp in length
          *    43691    43790: gap of 100 bp
          *    43791    47621: contig of 3831 bp in length
          *    47622    47721: gap of 100 bp
          *    47722    53852: contig of 6131 bp in length
          *    53853    53952: gap of 100 bp
          *    53953    63422: contig of 9470 bp in length
          *    63423    63522: gap of 100 bp
          *    63523    77361: contig of 13839 bp in length
          *    77362    77461: gap of 100 bp
          *    77462    92465: contig of 15004 bp in length
          *    92466    92565: gap of 100 bp
          *    92566   106951: contig of 14386 bp in length
          *   106952   107051: gap of 100 bp
          *   107052   129678: contig of 22627 bp in length
          *   129679   129778: gap of 100 bp
          *   129779   151427: contig of 21649 bp in length
          *   151428   151527: gap of 100 bp
          *   151528   181838: contig of 30311 bp in length
          *   181839   181938: gap of 100 bp
          *   181939   212120: contig of 30182 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212120
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-767B5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1043
                     /note="assembly_fragment"
     misc_feature    1144. .2224
                     /note="assembly_fragment"
     misc_feature    2325. .3345
                     /note="assembly_fragment"
     misc_feature    3446. .4502
                     /note="assembly_fragment"
     misc_feature    4603. .5756
                     /note="assembly_fragment"
     misc_feature    5857. .7097
                     /note="assembly_fragment"
     misc_feature    7198. .8236
                     /note="assembly_fragment"
     misc_feature    8337. .9353
                     /note="assembly_fragment"
     misc_feature    9454. .10705
                     /note="assembly_fragment"
     misc_feature    10806. .12332
                     /note="assembly_fragment"
     misc_feature    12433. .13877
                     /note="assembly_fragment"
     misc_feature    13978. .15201
                     /note="assembly_fragment"
     misc_feature    15302. .16968
                     /note="assembly_fragment"
     misc_feature    17069. .19060
                     /note="assembly_fragment"
     misc_feature    19161. .22639
                     /note="assembly_fragment"
     misc_feature    22740. .26376
                     /note="assembly_fragment"
     misc_feature    26477. .29275
                     /note="assembly_fragment"
     misc_feature    29376. .33265
                     /note="assembly_fragment"
     misc_feature    33366. .37012
                     /note="assembly_fragment"
     misc_feature    37113. .39390
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
```

```
     misc_feature    39491. .43690
                     /note="assembly_fragment"
     misc_feature    43791. .47621
                     /note="assembly_fragment"
     misc_feature    47722. .53852
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    53953. .63422
                     /note="assembly_fragment"
     misc_feature    63523. .77361
                     /note="assembly_fragment"
     misc_feature    77462. .92465
                     /note="assembly_fragment"
     misc_feature    92566. .106951
                     /note="assembly_fragment"
     misc_feature    107052. .129678

  Query Match             81.1%;  Score 808;  DB 2;  Length 212120;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  858;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180416 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 180475

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180476 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 180535

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180536 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180595

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     180596 GTAGCTTGACATTTTTCCATGTTAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 180655

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180656 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 180715

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180716 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 180775

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180776 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 180835

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180836 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 180895

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180896 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 180955

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180956 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 181015

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181016 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 181075

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181076 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 181135

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181136 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 181195

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181196 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 181255

Qy        841 TAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||
Db     181256 TAACAGTTATAAAGAAAGA 181274


RESULT 3
AX817316
LOCUS       AX817316                2463 bp    DNA     linear   PAT 10-DEC-2003
DEFINITION  Sequence 64 from Patent WO02081517.
ACCESSION   AX817316
VERSION     AX817316.1  GI:39722703
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Decristofaro,M.F., Padigaru,M., Miller,C., Tchernev,V., Zhong,H.,
            Zhong,M., Anderson,D., Ballinger,R., Gerlach,V., Spytek,K.A.,
            Rastelli,L., Kekuda,R., Guo,X., Zerhusen,B., Andrew,D., Mezes,P.,
            Patturajan,M., Burgess,C.E., Eisen,A., Wolenc,A., Baumgartner,J.,
            Shimkets,R.A., Gusev,V., Vernet,C.A., Taupier,R.J., Pena,C.,
            Shenoy,S., Li,L., Casman,S., Bolgog,F., Fernandes,E., Smithson,G.,
            Malyankar,U., Taillon,B. and Liu,X.
  TITLE     Novel polypeptides and nucleic acids encoded thereby
  JOURNAL   Patent: WO 02081517-A 64 17-OCT-2002;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2463
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             <3. .2108
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE84363.1"
                     /db_xref="GI:39722704"
                     /translation="WASRHTPAPSGGSAPSPARPRRLLRSAFSCELGRRRRRRSTFGE
                     PGGSGKLTPPLVPTLDRGALAAMREQLKGHETQTTCWDHPKMTELYQSLADLNNVRFS
                     AYRTAMKLRRLQKALCWDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLE
                     QEHNNLVNVPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRNLFK
                     QVASSTGFCDQRRLGLLLHDSIQIPRQLGEVASPGGSNIEPSVRSCFQFANNKPEIEA
                     ALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDI
                     CQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGY
                     LPVQTVLEGDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNG
                     SYLNDSISPNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERIL
                     ADLEEENRNLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQH
                     KGRLEARMQILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQP
                     MLLRVVGSQTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADD
                     LGRAMESLVSVMTDEEGAE"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 2463;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2009 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 2068

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2069 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 2128

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2129 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 2188
```

```
Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2189 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 2248

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2249 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 2308

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2309 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 2368

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2369 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 2418
```

RESULT 4
BC028720
LOCUS BC028720 4658 bp mRNA linear PRI 25-AUG-2003
DEFINITION Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker types), transcript variant Dp71b, mRNA (cDNA clone IMAGE:4822807), complete cds.
ACCESSION BC028720
VERSION BC028720.1 GI:20379675
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4658)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 4658)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (29-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 46 Row: j Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5032296
This clone has the following problem: The cds is short compared to the longest cds in the locus.

FEATURES Location/Qualifiers
source 1. .4658
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4822807"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
gene 1. .4658
/gene="DMD"
/note="synonyms: BMD, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272"
/db_xref="LocusID:1756"
/db_xref="MIM:300377"
CDS 130. .2037
/codon_start=1
/product="dystrophin, Dp71b isoform"
/protein_id="AAH28720.1"
/db_xref="GI:20379676"
/db_xref="LocusID:1756"
/translation="MREQLKGHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKL
RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVN
VPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGF
CDQRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMR
LEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSG
RVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLE
GDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSIS
PNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENR
NLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARM
QILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGS
QTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADDLGRAMESL
VSVMTDEEGAE"
misc_feature 844. .981
/note="ZZ; Region: Zinc finger, ZZ type. Zinc finger present in dystrophin, CBP/p300. ZZ in dystrophin binds calmodulin Putative zinc finger"
/db_xref="CDD:pfam00569"

ORIGIN

Query Match 41.2%; Score 410; DB 9; Length 4658;
Best Local Similarity 100.0%; Pred. No. 5.3e-203;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 1997

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 2057

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 2117

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 2177

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 2237
```

```
Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2238 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 2297

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2298 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 2347


RESULT 5
AX538620
LOCUS       AX538620                5339 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 40 from Patent WO0229056.
ACCESSION   AX538620
VERSION     AX538620.1  GI:25271166
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 40 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .5339
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 5339;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4551 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4610

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4611 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4730

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4731 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4790

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4791 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4850

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4851 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 4900


RESULT 6
AX538619
LOCUS       AX538619                5417 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 39 from Patent WO0229056.
ACCESSION   AX538619
VERSION     AX538619.1  GI:25271163
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 39 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .5417
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 5417;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4628

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4629 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4688

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4689 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4748

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4749 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4808

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4868

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4869 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4928

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 4978


RESULT 7
AX538621
LOCUS       AX538621                5462 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 41 from Patent WO0229056.
ACCESSION   AX538621
VERSION     AX538621.1  GI:25271168
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 41 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .5462
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 5462;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-203;
```

```
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4673

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4733

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4734 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4793

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4794 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4853

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4854 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4913

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4914 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4973

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       4974 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 5023


RESULT 8
AX538622
LOCUS       AX538622               8689 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 42 from Patent WO0229056.
ACCESSION   AX538622
VERSION     AX538622.1  GI:25271171
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 42 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .8689
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 8689;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 6044

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 6104

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6105 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 6164

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6165 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 6224

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6225 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 6284

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6285 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 6344

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       6345 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 6394


RESULT 9
AX538624
LOCUS       AX538624               11443 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 44 from Patent WO0229056.
ACCESSION   AX538624
VERSION     AX538624.1  GI:25271175
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 44 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .11443
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8739 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 8798

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 8858

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8859 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 8918

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8919 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 8978

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8979 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 9038

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9039 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 9098

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       9099 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 9148


RESULT 10
AX538627
LOCUS       AX538627               12057 bp    DNA     linear   PAT 23-NOV-2002
```

```
DEFINITION  Sequence 47 from Patent WO0229056.
ACCESSION   AX538627
VERSION     AX538627.1  GI:25271181
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chamberlain,J.S. and Harper,S.Q.
  TITLE     Mini-dystrophin nucleic acid and peptide sequences
  JOURNAL   Patent: WO 0229056-A 47 11-APR-2002;
            THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES             Location/Qualifiers
     source          1. .12057
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
ORIGIN

  Query Match             41.2%;  Score 410;  DB 6;  Length 12057;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11214 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11273

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11274 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11333

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11334 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11393

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11394 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11453

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11454 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11513

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11514 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11573

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11574 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11623


RESULT 11
HSDMDR
LOCUS       HSDMDR                  12446 bp    mRNA    linear   PRI 12-SEP-1993
DEFINITION  Human mRNA for dystrophin.
ACCESSION   X14298
VERSION     X14298.1  GI:30845
KEYWORDS    Dmd gene; Duchenne muscular dystrophy; dystrophin.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 12446)
  AUTHORS   Rosenthal,A., Speer,A., Billwitz,H., Cross,G.S., Forrest,S.M. and
            Davies,K.E.
  TITLE     Two human cDNA molecules coding for the Duchenne muscular dystrophy
            (DMD) locus are highly homologous
  JOURNAL   Nucleic Acids Res. 17 (13), 5391 (1989)
  MEDLINE   89345106
   PUBMED   2668885
REFERENCE   2  (bases 1 to 12446)
  AUTHORS   Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1989) Rosenthal A., Akademie der Wissenschaften
            der DDr, Zentralinstitut fuer Molekularbiologie, Robert-Roessle
            Str.10, 1115 Berlin Buch, DDR
COMMENT     see also M18533 and M20250 for Dmd seqs.; discrepancies compared to
            M18533 cDNA were located at x14298 pos. 496, 1772, 1965, 2449,
            3687, 4229, 4504, 5075, 5332, 5630 and 7194.
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X chromosomal, Xp21."
                     /tissue_type="muscle"
                     /dev_stage="fetal and adult."
     CDS             99. .11156
                     /note="unnamed protein product; dystrophin (AA 1 - 3685)"
                     /codon_start=1
                     /protein_id="CAA32479.1"
                     /db_xref="GI:30846"
                     /db_xref="GOA:P11532"
                     /db_xref="SWISS-PROT:P11532"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQPTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDV
                     DTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQ
                     QITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL
                     MESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTAH
                     QGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLM
                     DLQNQKLKELNDWLTKTEERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVN
                     SLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTE
                     EQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSIK
                     QDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVM
                     ETVTTVTTREQILVKHAQEELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEA
                     VLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSGQALVEQMVNEGVNA
                     DSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQ
                     PTTPSEPTAIKSQLKICKDEVNRLSGLQPQIERLKIQSIALKEKGQGPMFLDADFVAF
                     TNHFKQVFSDVQAREKELQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDY
                     EIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWK
                     KLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL
                     KQCRLLVSDIQTIQPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVY
                     ARKEALKGGLEKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQKAFEEMKRAK
                     EEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNGKCKTLE
                     EVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNP
                     NQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETENS
                     LHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEA
                     AQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKSVE
                     QEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHY
                     NELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSE
                     VAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIAVTSRAE
                     EWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELND
                     IRPKVDSTRDQAANLMANHGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELE
                     QFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQRGDNLQQRITDER
                     KSEEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKL
                     ASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWRE
                     IESKFAQFRRLNFAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQ
                     LLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSGRIDIIHSKKTAALQSATPVERVKL
                     QEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI
                     PENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNL
                     RWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKE
                     KLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKLKQTNLQWIKVS
                     RALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFD
                     VKETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRA
                     KQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWT
                     ELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLEELITAAQNLK
                     NKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVL
                     GQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSAD
                     DTRKVHMITENINASWRSIHKRVSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTA
                     NVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQKILRSLEGSD
                     DAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDEL
                     SRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQ
```

EPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNLHSADWQRKIDETLERLQELQEATD
ELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQL
TTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP
WERAISPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKA
LCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVD
MCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRL
GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSM
VWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGH
KMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
PVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESID
DEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEY
DRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDH
NKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSDSM
GEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM"

ORIGIN

Query Match 41.2%; Score 410; DB 9; Length 12446;
Best Local Similarity 100.0%; Pred. No. 4.7e-203;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11143 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11202

Qy      510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11203 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11262

Qy      570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11263 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11322

Qy      630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11323 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11382

Qy      690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11383 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11442

Qy      750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11443 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11502

Qy      810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11503 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11552
```

RESULT 12
AX409637
LOCUS AX409637 13957 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2284 from Patent WO0229103.
ACCESSION AX409637
VERSION AX409637.1 GI:21442342
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2284 11-APR-2002;
GENE LOGIC INC (US)
FEATURES Location/Qualifiers
source 1. .13957
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. M18533"

ORIGIN

Query Match 41.2%; Score 410; DB 6; Length 13957;
Best Local Similarity 100.0%; Pred. No. 4.6e-203;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy      510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy      570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy      630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy      690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy      750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy      810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662
```

RESULT 13
AX538581
LOCUS AX538581 13957 bp DNA linear PAT 23-NOV-2002
DEFINITION Sequence 1 from Patent WO0229056.
ACCESSION AX538581
VERSION AX538581.1 GI:25271086
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Chamberlain,J.S. and Harper,S.Q.
TITLE Mini-dystrophin nucleic acid and peptide sequences
JOURNAL Patent: WO 0229056-A 1 11-APR-2002;
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES Location/Qualifiers
source 1. .13957
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 41.2%; Score 410; DB 6; Length 13957;
Best Local Similarity 100.0%; Pred. No. 4.6e-203;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy      510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy      570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy      630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662


RESULT 14
HUMDYS
LOCUS       HUMDYS                 13957 bp    mRNA    linear   PRI 25-MAY-2000
DEFINITION  Homo sapiens dystrophin (DMD) mRNA, complete cds.
ACCESSION   M18533 M17154 M18026 M20250
VERSION     M18533.1  GI:181856
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1699)
  AUTHORS   Koenig,M., Hoffman,E.P., Bertelson,C.J., Monaco,A.P., Feener,C. and
            Kunkel,L.M.
  TITLE     Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
            preliminary genomic organization of the DMD gene in normal and
            affected individuals
  JOURNAL   Cell 50 (3), 509-517 (1987)
  MEDLINE   87273512
   PUBMED   3607877
REFERENCE   2  (bases 1678 to 3830)
  AUTHORS   Hoffman,E.P., Monaco,A.P., Feener,C.C. and Kunkel,L.M.
  TITLE     Conservation of the Duchenne muscular dystrophy gene in mice and
            humans
  JOURNAL   Science 238 (4825), 347-350 (1987)
  MEDLINE   88018015
   PUBMED   3659917
REFERENCE   3  (bases 1 to 13957)
  AUTHORS   Koenig,M., Monaco,A.P. and Kunkel,L.M.
  TITLE     The complete sequence of dystrophin predicts a rod-shaped
            cytoskeletal protein
  JOURNAL   Cell 53 (2), 219-226 (1988)
  MEDLINE   88194521
   PUBMED   3282674
COMMENT     On May 25, 2000 this sequence version replaced gi:340693.
            Draft entry and computer-readable sequence kindly provided by
            M.Koenig, 01-APR-1988 The severity of muscular dystrophy is
            determined by the size of the deleted DNA segment. Deletions found
            in different patients were from positions 302-2200, 473-1168,
            1691-1810, and 1169-3011.
FEATURES             Location/Qualifiers
     source          1. .13957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp21.3-p21.1"
                     /tissue_type="muscle"
                     /dev_stage="fetus"
                     /note="G00-119-850"
     gene            1. .13957
                     /gene="DMD"
     CDS             209. .11266
                     /gene="DMD"
                     /codon_start=1
                     /product="dystrophin"
                     /protein_id="AAA53189.1"
                     /db_xref="GI:181857"
                     /translation="MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDL
                     QDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVDG
                     NHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT
                     TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDV
                     DTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQ
                     QITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL
                     MESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTAH
                     QGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLM
                     DLQNQKLKELNDWLTKTEERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVN
                     SLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTE
                     EQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSLK
                     QDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVM
                     ETVTTVTTREQILVKHAQEELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEA
                     VLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRSAQALVEQMVNEGVNA
                     DSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQ
                     PTTPSEPTAIKSQLKICKDEVNRLSGLQPQIERLKIQSIALKEKGQGPMFLDADFVAF
                     TNHFKQVPSDVQAREKELQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDY
                     EIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWK
                     KLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL
                     KQCRLLVSDIQTIQPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVY
                     ARKEALKGGLEKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTPDELQKAVEEMKRAK
                     EEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNGKCKTLE
                     EVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNP
                     NQIRILAQTLTDGGVMDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETEKS
                     LHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEA
                     AQRVLSQIDVAQKKLQDVSMKFRLFQKPANFELRLQESKMILDEVKMHLPALETKSVE
                     QEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHY
                     NELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSE
                     VAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIAVTSRAE
                     EWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELND
                     IRPKVDSTRDQAANLMANRGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELE
                     QFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQRGDNLQQRITDER
                     KREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKL
                     ASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWRE
                     IESKFAQFRRLNFAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQ
                     LLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSGRIDIIHSKKTAALQSATPVERVKL
                     QEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI
                     PENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNL
                     RWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKE
                     KLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKLKQTNLQWIKVS
                     RALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFD
                     VQETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRA
                     KQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWT
                     ELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLEELITAAQNLK
                     NKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVL
                     GQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSAD
                     DTRKVHMITENINASWRSIHKRVSEREAALEETHRLLQQFPLDLEKFLAWLTEAETTA
                     NVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQKILRSLEGSD
                     DAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDEL
                     SRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQ
                     EPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNLHSADWQRKIDETLERLQELQEATD
                     ELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQL
                     TTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP
                     WERAISPNKVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKA
                     LCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVD
                     MCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRL
                     GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSM
                     VWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGH
                     KMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
                     PVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESID
                     DEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEY
                     DRLKQQHEHKGLSPLPSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDH
                     NKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGSQTSDSM
                     GEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM"
ORIGIN

  Query Match           41.2%;  Score 410;  DB 9;  Length 13957;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-203;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662
```

RESULT 15
G15848
LOCUS G15848 2148 bp DNA linear STS 19-JAN-1996
DEFINITION human STS CHLC.UTR_01924_M18533.P56108 clone UTR_01924_M18533, sequence tagged site.
ACCESSION G15848
VERSION G15848.1 GI:1161737
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2148)
AUTHORS Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
TITLE Cooperative Human Linkage Center
JOURNAL Unpublished (1995)
COMMENT Synonyms: UTR_01924_M18533, CHLC.UTR_01924_M18533.T36152
Contact: Dr. Jeffrey C. Murray
UofI
The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jeff-murray@uiowa.edu

Primer A: AACGCATTTTGGGTTGTTTA
Primer B: GATATCACGCCAAAAGGATG
STS size: 189
PCR Profile:

denature:	30 seconds at 94 degrees C
annealing:	75 seconds at 55 degrees C
extension:	15 seconds at 72 degrees C
PCR cycles:	27
extension:	6 minutes at 72 degress C

Protocol:

Template:	30ng genomic DNA
Primer:	each 1.5 pmole
dNTPs:	each 200 uM
Taq Polymerase:	0.3 units
Total Vol:	10 ul

Buffer:

MgCl2:	1.5mM
KCl:	50mM
Tris:	10mM
pH:	8.3

Prepared with primer pairs derived from M18533.

```
FEATURES             Location/Qualifiers
     source          1. .2148
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     STS             350. .538
     primer_bind     350. .369
     primer_bind     complement(519. .538)
ORIGIN
```

Query Match 39.8%; Score 396; DB 11; Length 2148;
Best Local Similarity 100.0%; Pred. No. 1.3e-195;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        464 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 60

Qy        524 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 120

Qy        584 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 180

Qy        644 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 240

Qy        704 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 300

Qy        764 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 360

Qy        824 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||
Db        361 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 396
```

Search completed: April 6, 2004, 16:30:24
Job time : 4178.52 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:14:32 ; Search time 109.885 Seconds
(without alignments)
5030.075 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	359	36.0	13977	4	US-09-484-970B-60	Sequence 60, Appl
c	2	126	12.7	13977	4	US-09-484-970B-60	Sequence 60, Appl
c	3	91	9.1	19307	3	US-08-836-022A-10	Sequence 10, Appl
c	4	91	9.1	19307	3	US-09-427-048A-10	Sequence 10, Appl
	5	42	4.2	19307	3	US-08-836-022A-10	Sequence 10, Appl
	6	42	4.2	19307	3	US-09-427-048A-10	Sequence 10, Appl
	7	23	2.3	434	3	US-08-943-731-127	Sequence 127, App
	8	23	2.3	24183	3	US-08-943-731-3	Sequence 3, Appli
	9	20	2.0	891	4	US-08-936-165A-90	Sequence 90, Appl
c	10	20	2.0	1201	3	US-09-274-642-1	Sequence 1, Appli
	11	19	1.9	4766	5	PCT-US93-07261-10	Sequence 10, Appl
c	12	19	1.9	9139	4	US-09-322-478-22	Sequence 22, Appl
c	13	18	1.8	25	2	US-08-676-782-4	Sequence 4, Appli
c	14	18	1.8	342	2	US-08-676-782-11	Sequence 11, Appl
c	15	18	1.8	393	1	US-08-248-505-1	Sequence 1, Appli
c	16	18	1.8	393	2	US-08-676-782-6	Sequence 6, Appli
	17	18	1.8	750	4	US-09-107-532A-866	Sequence 866, App
	18	18	1.8	889	2	US-08-935-886-7	Sequence 7, Appli
	19	18	1.8	889	2	US-08-935-886-13	Sequence 13, Appl
	20	18	1.8	1098	4	US-08-956-171E-253	Sequence 253, App
c	21	18	1.8	1349	2	US-08-676-782-10	Sequence 10, Appl
	22	18	1.8	1824	4	US-09-489-039A-4969	Sequence 4969, Ap
	23	18	1.8	1919	4	US-09-398-179-1	Sequence 1, Appli
c	24	18	1.8	1919	4	US-09-398-179-2	Sequence 2, Appli
c	25	18	1.8	1966	4	US-09-221-017B-964	Sequence 964, App
	26	18	1.8	2076	4	US-09-134-001C-1838	Sequence 1838, Ap
	27	18	1.8	2382	4	US-09-107-532A-2569	Sequence 2569, Ap
	28	18	1.8	2415	4	US-09-107-532A-2694	Sequence 2694, Ap
c	29	17	1.7	92	1	US-07-839-751-1	Sequence 1, Appli
c	30	17	1.7	92	1	US-08-239-427A-1	Sequence 1, Appli
c	31	17	1.7	92	2	US-08-467-975-1	Sequence 1, Appli
c	32	17	1.7	92	3	US-08-401-355-1	Sequence 1, Appli
	33	17	1.7	305	4	US-09-313-294A-7590	Sequence 7590, Ap
	34	17	1.7	391	4	US-09-621-976-11175	Sequence 11175, A
c	35	17	1.7	483	4	US-09-107-532A-498	Sequence 498, App
c	36	17	1.7	757	4	US-09-023-655-769	Sequence 769, App
	37	17	1.7	888	4	US-09-134-000C-907	Sequence 907, App
	38	17	1.7	1079	4	US-09-976-594-780	Sequence 780, App
c	39	17	1.7	1260	3	US-09-081-689-3	Sequence 3, Appli
c	40	17	1.7	1297	4	US-08-858-207A-66	Sequence 66, Appl
c	41	17	1.7	1329	4	US-09-305-984-13	Sequence 13, Appl
c	42	17	1.7	1329	4	US-09-073-541A-13	Sequence 13, Appl
c	43	17	1.7	1329	4	US-09-493-940-13	Sequence 13, Appl
	44	17	1.7	1332	3	US-09-457-046B-49	Sequence 49, Appl
c	45	17	1.7	1374	4	US-09-134-001C-1447	Sequence 1447, Ap

ALIGNMENTS

RESULT 1
US-09-484-970B-60
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
; NAME/KEY: unsure
; LOCATION: 11721-11761, 12294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match 36.0%; Score 359; DB 4; Length 13977;
Best Local Similarity 99.8%; Pred. No. 6.3e-170;
Matches 409; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11254 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11313

Qy      510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11314 CTTAGTATCATTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11373

Qy      570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11374 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11433

Qy      630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11434 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11493

Qy      690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11494 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11553
```

```
Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11554 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11613

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11614 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11663


RESULT 2
US-09-484-970B-60/c
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Karen A.
;  APPLICANT:  Volkmuth, Wayne
;  APPLICANT:  Walker, Michael G.
;  TITLE OF INVENTION:  BONE REMODELING GENES
;  FILE REFERENCE:  PB-0014 US
;  CURRENT APPLICATION NUMBER: US/09/484,970B
;  CURRENT FILING DATE:  2000-01-18
;  NUMBER OF SEQ ID NOS:  172
;  SOFTWARE:  PERL Program
; SEQ ID NO 60
;    LENGTH: 13977
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
;    NAME/KEY: unsure
;    LOCATION: 11721-11761, 12294, 13969
;    OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

  Query Match             12.7%;  Score 126;  DB 4;  Length 13977;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-53;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 AAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13315 AAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA 13256

Qy        911 CTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13255 CTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 13196

Qy        971 GAAGCC 976
              ||||||
Db      13195 GAAGCC 13190


RESULT 3
US-08-836-022A-10/c
; Sequence 10, Application US/08836022A
; Patent No. 6001557
;  GENERAL INFORMATION:
;    APPLICANT:  Trustees of the University of Pennsylvania
;    APPLICANT:  Wilson, James M.
;    APPLICANT:  Fisher, Krishna J.
;    APPLICANT:  Chen, Shu-Jen
;    APPLICANT:  Weitzman, Matthew
;    TITLE OF INVENTION:  Improved Adenovirus Virus and
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P O Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,022A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/331,381
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GNVPN.008PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19307 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-836-022A-10

  Query Match             9.1%;  Score 91;  DB 3;  Length 19307;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-36;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        665 TATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTT 724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3225 TATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTT 3166

Qy        725 TACACGTCTATGCAATTGTACAAAAAAGTTA 755
              |||||||||||||||||||||||||||||||
Db       3165 TACACGTCTATGCAATTGTACAAAAAAGTTA 3135


RESULT 4
US-09-427-048A-10/c
; Sequence 10, Application US/09427048A
; Patent No. 6203975
;  GENERAL INFORMATION:
;       APPLICANT: Trustees of the University of Pennsylvania
;                  Wilson, James M.
;                  Fisher, Krishna J.
;                  Chen, Shu-Jen
;                  Weitzman, Matthew
;       TITLE OF INVENTION: Improved Adenovirus Virus and
;                           Methods of Use Thereof
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Cntr, P O Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/427,048A
;            FILING DATE: 21-Oct-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/836,022
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
```

```
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 19307 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

  Query Match             9.1%;  Score 91;  DB 3;  Length 19307;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-36;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        665 TATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTT 724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3225 TATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTT 3166

Qy        725 TACACGTCTATGCAATTGTACAAAAAAGTTA 755
              |||||||||||||||||||||||||||||||
Db       3165 TACACGTCTATGCAATTGTACAAAAAAGTTA 3135


RESULT 5
US-08-836-022A-10
; Sequence 10, Application US/08836022A
; Patent No. 6001557
;  GENERAL INFORMATION:
;    APPLICANT:  Trustees of the University of Pennsylvania
;    APPLICANT:  Wilson, James M.
;    APPLICANT:  Fisher, Krishna J.
;    APPLICANT:  Chen, Shu-Jen
;    APPLICANT:  Weitzman, Matthew
;    TITLE OF INVENTION:  Improved Adenovirus Virus and
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P O Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,022A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/331,381
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GNVPN.008PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19307 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-836-022A-10

  Query Match             4.2%;  Score 42;  DB 3;  Length 19307;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-11;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 975
              |||||||||||||||||||||||||||||||||||||||||||
Db       1563 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 1604


RESULT 6
US-09-427-048A-10
; Sequence 10, Application US/09427048A
; Patent No. 6203975
;  GENERAL INFORMATION:
;       APPLICANT: Trustees of the University of Pennsylvania
;                  Wilson, James M.
;                  Fisher, Krishna J.
;                  Chen, Shu-Jen
;                  Weitzman, Matthew
;       TITLE OF INVENTION: Improved Adenovirus Virus and
;                           Methods of Use Thereof
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Cntr, P O Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/427,048A
;            FILING DATE: 21-Oct-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/836,022
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 19307 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

  Query Match             4.2%;  Score 42;  DB 3;  Length 19307;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-11;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        934 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 975
              |||||||||||||||||||||||||||||||||||||||||||
Db       1563 AACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGAGAAGC 1604


RESULT 7
US-08-943-731-127
; Sequence 127, Application US/08943731
; Patent No. 6265157
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  PROCKOP, DARWIN J.
;    APPLICANT:  SPOTILA, LORETTA D.
;    APPLICANT:  DELTAS, CONSTANTINOS D.
;    APPLICANT:  SEREDA, LARISA
;    APPLICANT:  LARSON, ANDREA W.
;    APPLICANT:  PACK, MICHAEL
;    APPLICANT:  COLIGE, ALAIN
;    APPLICANT:  EARLY, JAMES
;    APPLICANT:  KORKKO, JARMO
;    APPLICANT:  ALA-KOKKO, LEENA, et al.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;    NUMBER OF SEQUENCES:  666
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;      STREET:  FLR.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,731
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,322
;      FILING DATE:  14-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/803,628
;      FILING DATE:  03-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9598-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  127:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  434 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-943-731-127

  Query Match             2.3%;  Score 23;  DB 3;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 0.084;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        423 GTCTGTCTTTCTTTCTCTTTGTT 445
              |||||||||||||||||||||||
Db        395 GTCTGTCTTTCTTTCTCTTTGTT 417


RESULT 8
US-08-943-731-3
; Sequence 3, Application US/08943731
; Patent No. 6265157
;  GENERAL INFORMATION:
;    APPLICANT:  PROCKOP, DARWIN J.
;    APPLICANT:  SPOTILA, LORETTA D.
;    APPLICANT:  DELTAS, CONSTANTINOS D.
;    APPLICANT:  SEREDA, LARISA
;    APPLICANT:  LARSON, ANDREA W.
;    APPLICANT:  PACK, MICHAEL
;    APPLICANT:  COLIGE, ALAIN
;    APPLICANT:  EARLY, JAMES
;    APPLICANT:  KORKKO, JARMO
;    APPLICANT:  ALA-KOKKO, LEENA, et al.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;    NUMBER OF SEQUENCES:  666
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;      STREET:  FLR.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,731
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,322
;      FILING DATE:  14-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/803,628
;      FILING DATE:  03-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9598-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-943-731-3

  Query Match             2.3%;  Score 23;  DB 3;  Length 24183;
  Best Local Similarity   100.0%;  Pred. No. 0.09;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        423 GTCTGTCTTTCTTTCTCTTTGTT 445
              |||||||||||||||||||||||
Db      10136 GTCTGTCTTTCTTTCTCTTTGTT 10158


RESULT 9
US-08-936-165A-90
; Sequence 90, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
;    APPLICANT:  Ward, Judith
;    TITLE OF INVENTION:  No. 6348582el Prokaryotic Polynucleotides,
```

```
;    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  90:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  891 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-936-165A-90

  Query Match             2.0%;  Score 20;  DB 4;  Length 891;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        686 TTCTAAGTCTGTTATTGTTT 705
              ||||||||||||||||||||
Db        579 TTCTAAGTCTGTTATTGTTT 598


RESULT 10
US-09-274-642-1/c
; Sequence 1, Application US/09274642A
; Patent No. 6071729
; GENERAL INFORMATION:
;  APPLICANT: Jeffries, Thomas W.
;  APPLICANT: Shi, Nian-Qing
;  TITLE OF INVENTION: Disruption of cytochrome c gene in xylose-fermenting yeast
;  FILE REFERENCE: cytochrome c
;  CURRENT APPLICATION NUMBER: US/09/274,642A
;  CURRENT FILING DATE: 1999-03-23
;  EARLIER APPLICATION NUMBER: 60/080,493
;  EARLIER FILING DATE: 1998-04-02
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1201
;   TYPE: DNA
;   ORGANISM: Pichia stipitis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (610)..(942)
US-09-274-642-1

  Query Match             2.0%;  Score 20;  DB 3;  Length 1201;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        623 TTACAAGAAATAAATCTATA 642
              ||||||||||||||||||||
Db       1062 TTACAAGAAATAAATCTATA 1043


RESULT 11
PCT-US93-07261-10
; Sequence 10, Application PC/TUS9307261
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John H. C. Blasdale
;      STREET:  One Giralda Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07940-1000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.5
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07261
;      FILING DATE:  19930805
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/927,531
;      FILING DATE:  07-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Blasdale, John H. C.
;      REGISTRATION NUMBER:  31,895
;      REFERENCE/DOCKET NUMBER:  DX0288K
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-822-7398
;      TELEFAX:  201-822-7039
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4766 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium falciparum
;      STRAIN:  Malayan Camp
;    IMMEDIATE SOURCE:
;      CLONE:  p2b1;p12-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..4766
PCT-US93-07261-10

  Query Match             1.9%;  Score 19;  DB 5;  Length 4766;
  Best Local Similarity   100.0%;  Pred. No. 8.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        864 TAAAGGAAAAAGAAAATAA 882
              |||||||||||||||||||
Db        190 TAAAGGAAAAAGAAAATAA 208


RESULT 12
US-09-322-478-22/c
; Sequence 22, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
;  APPLICANT: Wright, David A.
;  APPLICANT: Voytas, Daniel F.
;  TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
;  FILE REFERENCE: P-1065 ISURF Plant Retroelement
```

```
;  CURRENT APPLICATION NUMBER: US/09/322,478
;  CURRENT FILING DATE: 1999-05-28
;  EARLIER APPLICATION NUMBER: 60/087125
;  EARLIER FILING DATE: 1998-05-29
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 9139
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-322-478-22

  Query Match             1.9%;  Score 19;  DB 4;  Length 9139;
  Best Local Similarity   100.0%;  Pred. No. 8.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        528 AGATGAAGAAGGAGCAGAA 546
              ||||||||||||||||||||
Db       5801 AGATGAAGAAGGAGCAGAA 5783


RESULT 13
US-08-676-782-4/c
; Sequence 4, Application US/08676782
; Patent No. 5976792
;  GENERAL INFORMATION:
;    APPLICANT:  CHEUNG, Ambrose
;    APPLICANT:  FISCHETTI, Vincent A.
;    TITLE OF INVENTION:  REGULATION OF EXOPROTEIN IN
;    TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,782
;      FILING DATE:  08-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/248,505
;      FILING DATE:  25-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  016921-092
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-676-782-4

  Query Match             1.8%;  Score 18;  DB 2;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        431 TTCTTTCTCTTTGTTTTC 448
              ||||||||||||||||||
Db         24 TTCTTTCTCTTTGTTTTC 7


RESULT 14
US-08-676-782-11/c
; Sequence 11, Application US/08676782
; Patent No. 5976792
;  GENERAL INFORMATION:
;    APPLICANT:  CHEUNG, Ambrose
;    APPLICANT:  FISCHETTI, Vincent A.
;    TITLE OF INVENTION:  REGULATION OF EXOPROTEIN IN
;    TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,782
;      FILING DATE:  08-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/248,505
;      FILING DATE:  25-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  016921-092
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  157
;      OTHER INFORMATION:  /note= "Nucleotide at position 157
;      OTHER INFORMATION:  is N wherein N = C or T."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  340
;      OTHER INFORMATION:  /note= "Nucleotide at position 340
;      OTHER INFORMATION:  is N wherein N = C or T."
US-08-676-782-11

  Query Match             1.8%;  Score 18;  DB 2;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        431 TTCTTTCTCTTTGTTTTC 448
              ||||||||||||||||||
Db        150 TTCTTTCTCTTTGTTTTC 133


RESULT 15
US-08-248-505-1/c
; Sequence 1, Application US/08248505
; Patent No. 5587288
;  GENERAL INFORMATION:
;    APPLICANT:  CHEUNG, Ambrose
;    APPLICANT:  FISHETTI, Vincent A
```

```
;    TITLE OF INVENTION:  REGULATION OF EXOPROTEIN IN
;    TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/248,505
;      FILING DATE:  24-MAY-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  016921-018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..390
US-08-248-505-1

  Query Match             1.8%;  Score 18;  DB 1;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        431 TTCTTTCTCTTTGTTTTC 448
              ||||||||||||||||||
Db        168 TTCTTTCTCTTTGTTTTC 151


Search completed: April  6, 2004, 17:42:28
Job time : 111.385 secs
```

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 15:01:53 ; Search time 449.211 Seconds
(without alignments)
8309.593 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa..........gtgtttgatgttaattaatt 996

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2470632 seqs, 1873875610 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	996	100.0	996	9	US-09-966-264-2	Sequence 2, Appli
	2	410	41.2	5339	15	US-10-149-736-40	Sequence 40, Appl
	3	410	41.2	5417	15	US-10-149-736-39	Sequence 39, Appl
	4	410	41.2	5462	15	US-10-149-736-41	Sequence 41, Appl
	5	410	41.2	8689	15	US-10-149-736-42	Sequence 42, Appl
	6	410	41.2	11443	15	US-10-149-736-44	Sequence 44, Appl
	7	410	41.2	12057	15	US-10-149-736-47	Sequence 47, Appl
	8	410	41.2	13957	9	US-09-782-378A-22	Sequence 22, Appl
	9	410	41.2	13957	9	US-09-880-107-2284	Sequence 2284, Ap
	10	410	41.2	13957	15	US-10-149-736-1	Sequence 1, Appli
	11	410	41.2	14069	12	US-10-342-887-434	Sequence 434, App
	12	410	41.2	14082	12	US-10-342-887-981	Sequence 981, App
	13	410	41.2	14082	15	US-10-341-434-108	Sequence 108, App
	14	396	39.8	2691	15	US-10-149-736-38	Sequence 38, Appl
c	15	146	14.7	430	9	US-09-796-692-3505	Sequence 3505, Ap
c	16	146	14.7	430	14	US-10-040-862-3505	Sequence 3505, Ap
c	17	146	14.7	430	15	US-10-057-475B-3505	Sequence 3505, Ap
c	18	146	14.7	430	15	US-10-154-884B-3505	Sequence 3505, Ap
c	19	146	14.7	2691	15	US-10-149-736-38	Sequence 38, Appl
c	20	146	14.7	8689	15	US-10-149-736-42	Sequence 42, Appl
c	21	146	14.7	11443	15	US-10-149-736-44	Sequence 44, Appl
c	22	146	14.7	13957	9	US-09-782-378A-22	Sequence 22, Appl
c	23	146	14.7	13957	9	US-09-880-107-2284	Sequence 2284, Ap
c	24	146	14.7	13957	15	US-10-149-736-1	Sequence 1, Appli
c	25	146	14.7	14069	12	US-10-342-887-434	Sequence 434, App
c	26	146	14.7	14082	12	US-10-342-887-981	Sequence 981, App
c	27	146	14.7	14082	15	US-10-341-434-108	Sequence 108, App
	28	137	13.8	137	9	US-09-966-264-1	Sequence 1, Appli
	29	91	9.1	13815	15	US-10-149-736-2	Sequence 2, Appli
c	30	42	4.2	42	9	US-09-966-264-8	Sequence 8, Appli
c	31	42	4.2	13815	15	US-10-149-736-2	Sequence 2, Appli
	32	23	2.3	23	9	US-09-966-264-21	Sequence 21, Appl
	33	22	2.2	22	9	US-09-966-264-7	Sequence 7, Appli
c	34	22	2.2	22	9	US-09-966-264-12	Sequence 12, Appl
c	35	22	2.2	22	9	US-09-966-264-13	Sequence 13, Appl
c	36	22	2.2	22	9	US-09-966-264-29	Sequence 29, Appl
	37	21	2.1	21	9	US-09-966-264-10	Sequence 10, Appl
	38	20	2.0	20	9	US-09-966-264-9	Sequence 9, Appli
c	39	20	2.0	20	9	US-09-966-264-24	Sequence 24, Appl
c	40	20	2.0	36	15	US-10-149-736-65	Sequence 65, Appl
c	41	20	2.0	175	12	US-10-424-599-36603	Sequence 36603, A
c	42	20	2.0	459	12	US-10-424-599-53311	Sequence 53311, A
c	43	20	2.0	690	15	US-10-131-827-8211	Sequence 8211, Ap
c	44	20	2.0	734	12	US-10-424-599-45245	Sequence 45245, A
c	45	20	2.0	859	15	US-10-027-632-33432	Sequence 33432, A

ALIGNMENTS

RESULT 1
US-09-966-264-2
; Sequence 2, Application US/09966264
; Patent No. US20020099015A1
; GENERAL INFORMATION:
; APPLICANT: Barber, Elizabeth K
; TITLE OF INVENTION: Gene Expression Control Element DNA
; FILE REFERENCE: 896034605001
; CURRENT APPLICATION NUMBER: US/09/966,264
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/237,079
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 996
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(996)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (710)..(996)
; OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph
; OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a
; OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence
; OTHER INFORMATION: in the dystrophin DNA
US-09-966-264-2

Query Match 100.0%; Score 996; DB 9; Length 996;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
```

```
Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
```

RESULT 2
US-10-149-736-40
; Sequence 40, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 5339
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-40

Query Match 41.2%; Score 410; DB 15; Length 5339;
Best Local Similarity 100.0%; Pred. No. 2.6e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4551 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4610

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4611 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4730

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4731 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4790

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4791 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4850

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4851 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 4900
```

RESULT 3
US-10-149-736-39
; Sequence 39, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 5417
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic

US-10-149-736-39

Query Match 41.2%; Score 410; DB 15; Length 5417;
Best Local Similarity 100.0%; Pred. No. 2.6e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4628

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4629 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4688

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4689 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4748

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4749 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4808

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4809 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4868

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4869 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4928

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4929 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 4978
```

RESULT 4
US-10-149-736-41
; Sequence 41, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 5462
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-41

Query Match 41.2%; Score 410; DB 15; Length 5462;
Best Local Similarity 100.0%; Pred. No. 2.6e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4673

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4733

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4734 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4793

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4794 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4853

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4854 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4913

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4914 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4973

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4974 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 5023
```

RESULT 5
US-10-149-736-42
; Sequence 42, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 8689
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-42

Query Match 41.2%; Score 410; DB 15; Length 8689;
Best Local Similarity 100.0%; Pred. No. 2.6e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 6044

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 6104

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6105 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 6164

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6165 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 6224

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6225 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 6284

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6285 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 6344
```

```
Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6345 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 6394

RESULT 6
US-10-149-736-44
; Sequence 44, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
;   LENGTH: 11443
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-44

  Query Match             41.2%;  Score 410;  DB 15;  Length 11443;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8739 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 8798

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 8858

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8859 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 8918

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8919 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 8978

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8979 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 9038

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9039 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 9098

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9099 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 9148

RESULT 7
US-10-149-736-47
; Sequence 47, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
;   LENGTH: 12057
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-149-736-47

  Query Match             41.2%;  Score 410;  DB 15;  Length 12057;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11214 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11273

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11274 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11333

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11334 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11393

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11394 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11453

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11454 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11513

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11514 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11573

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11574 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11623

RESULT 8
US-09-782-378A-22
; Sequence 22, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
;  APPLICANT: Hearing, Patrick
;  APPLICANT:  Bahou, Wadie
;  APPLICANT:  Sandalon, Ziv
;  APPLICANT:  Gnatenko, Dmitri
;  TITLE OF INVENTION: Adenoviral Vectors
;  FILE REFERENCE: STONYB-04970
;  CURRENT APPLICATION NUMBER: US/09/782,378A
;  CURRENT FILING DATE:  2001-02-12
;  PRIOR APPLICATION NUMBER: 60/237,747
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 13957
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-782-378A-22

  Query Match             41.2%;  Score 410;  DB 9;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
```

```
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662


RESULT 9
US-09-880-107-2284
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
;   LENGTH: 13957
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284

  Query Match             41.2%;  Score 410;  DB 9;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662


RESULT 10
US-10-149-736-1
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 13957
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-149-736-1

  Query Match             41.2%;  Score 410;  DB 15;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662
```

```
RESULT 11
US-10-342-887-434
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
;   LENGTH: 14069
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-342-887-434

  Query Match             41.2%;  Score 410;  DB 12;  Length 14069;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11365 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11424

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11425 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11484

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11485 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11544

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11545 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11604

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11605 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11664

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11665 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11724

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11725 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11774


RESULT 12
US-10-342-887-981
; Sequence 981, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 981
;   LENGTH: 14082
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-342-887-981

  Query Match             41.2%;  Score 410;  DB 12;  Length 14082;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11378 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11437

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11438 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11497

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11498 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11557

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11558 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11617

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11618 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11677

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11678 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11737

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11738 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11787


RESULT 13
US-10-341-434-108
; Sequence 108, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 108
;   LENGTH: 14082
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (703)..(11388)
;   OTHER INFORMATION:
US-10-341-434-108

  Query Match             41.2%;  Score 410;  DB 15;  Length 14082;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-189;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11378 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11437

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11438 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11497

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11498 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11557

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11558 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11617

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11618 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11677

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11678 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11737

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11738 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11787


RESULT 14
US-10-149-736-38
; Sequence 38, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
;  APPLICANT: Chamberlain, Jeffrey S.
;  APPLICANT:  Harper, Scott Q.
;  TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
;  FILE REFERENCE: UM-06968
;  CURRENT APPLICATION NUMBER: US/10/149,736
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: PCT/US01/31126
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: 60/238,848
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
;   LENGTH: 2691
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-149-736-38

  Query Match             39.8%;  Score 396;  DB 15;  Length 2691;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-182;
  Matches  396;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        464 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 60

Qy        524 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 120

Qy        584 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 180

Qy        644 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 240

Qy        704 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 300

Qy        764 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 360

Qy        824 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||
Db        361 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 396


RESULT 15
US-09-796-692-3505/c
; Sequence 3505, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3505
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-796-692-3505

  Query Match             14.7%;  Score 146;  DB 9;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-60;
  Matches  146;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        851 AAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA 910
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 AAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA 237

Qy        911 CTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 970
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 CTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGAAATAGCATGA 177

Qy        971 GAAGCCGTGTTTGATGTTAATTAATT 996
                |||||||||||||||||||||||||
Db        176 GAAGCCGTGTTTGATGTTAATTAATT 151
```

Search completed: April 6, 2004, 17:51:13
Job time : 450.211 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:42:46 ; Search time 472.067 Seconds
(without alignments)
8963.144 Million cell updates/sec

Title: US-09-966-264D-2
Perfect score: 996
Sequence: 1 gtggtttgattgatagtaaa.....gtgtttgatgttaattaatt 996

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	996	100.0	996	6	ABK86497	Abk86497 Human Apo
2	996	100.0	1230	6	ABK86462	Abk86462 Human Apo
3	996	100.0	1234	6	ABK86463	Abk86463 Human Apo
4	410	41.2	2463	7	ABT33375	Abt33375 NOVX DNA
5	410	41.2	5339	6	ABK81998	Abk81998 DNA encod
6	410	41.2	5417	6	ABK81997	Abk81997 DNA encod
7	410	41.2	5462	6	ABK81999	Abk81999 DNA encod
8	410	41.2	8689	6	ABK82000	Abk82000 DNA encod
9	410	41.2	11443	6	ABK82002	Abk82002 DNA encod
10	410	41.2	12923	1	AAN90338	Aan90338 Sequence
11	410	41.2	13957	6	ABK81959	Abk81959 cDNA enco
12	410	41.2	13957	6	ABT10904	Abt10904 Human bre
13	410	41.2	13957	6	ABN95786	Abn95786 Gene #228
14	410	41.2	13957	6	ABS69900	Abs69900 Human dys
15	396	39.8	2691	6	ABK81996	Abk81996 Human dys
16	359	36.0	13977	6	ABS70403	Abs70403 Human bon
17	260	26.1	3163	3	AAZ48571	Aaz48571 A rod sho
18	260	26.1	3172	3	AAZ48570	Aaz48570 A rod sho
19	260	26.1	4075	3	AAZ48569	Aaz48569 A rod sho
20	260	26.1	4402	3	AAZ48568	Aaz48568 A rod sho
21	260	26.1	4402	3	AAZ48567	Aaz48567 A rod sho
22	251	25.2	3747	3	AAZ48566	Aaz48566 A rod sho
c 23	158	15.9	158	6	ABK86469	Abk86469 Human apo
c 24	146	14.7	200	6	ABK86468	Abk86468 Human dys
c 25	146	14.7	455	9	ADD32504	Add32504 Human mit
c 26	146	14.7	2691	6	ABK81996	Abk81996 Human dys
c 27	146	14.7	8689	6	ABK82000	Abk82000 DNA encod
c 28	146	14.7	11443	6	ABK82002	Abk82002 DNA encod
c 29	146	14.7	13957	6	ABK81959	Abk81959 cDNA enco
c 30	146	14.7	13957	6	ABT10904	Abt10904 Human bre
c 31	146	14.7	13957	6	ABN95786	Abn95786 Gene #228
c 32	146	14.7	13957	6	ABS69900	Abs69900 Human dys
33	145	14.6	2005	7	ABT33376	Abt33376 NOVX DNA
34	137	13.8	137	6	ABK86496	Abk86496 Human Apo
c 35	126	12.7	13977	6	ABS70403	Abs70403 Human bon
36	94	9.4	1044	6	ABK86464	Abk86464 Mouse gen
37	91	9.1	13815	2	AAV18885	Aav18885 Mus muscu
38	91	9.1	13815	6	ABK81960	Abk81960 cDNA enco
39	91	9.1	13815	6	ABI99799	Abi99799 Mouse isc
c 40	91	9.1	19307	2	AAT27558	Aat27558 Shuttle v
41	60	6.0	60	6	ABK86471	Abk86471 Human apo
c 42	46	4.6	959	6	ABN74601	Abn74601 Bovine em
c 43	42	4.2	42	6	ABK86474	Abk86474 Human apo
c 44	42	4.2	13815	2	AAV18885	Aav18885 Mus muscu
c 45	42	4.2	13815	6	ABK81960	Abk81960 cDNA enco

ALIGNMENTS

```
RESULT 1
ABK86497
ID   ABK86497 standard; cDNA; 996 BP.
XX
AC   ABK86497;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 cDNA.
XX
KW   Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .996
FT                   /*tag= a
FT                   /product= "Apo-dystrophin-4 protein appearing as
FT                   AAU98739"
FT                   /partial
FT                   /note= "No start or stop codon shown"
FT                   /transl_except= (pos:7. .9,aa:Xaa)
FT                   /transl_except= (pos:49. .51,aa:Xaa)
FT                   /transl_except= (pos:187. .189,aa:Xaa)
FT                   /transl_except= (pos:202. .204,aa:Xaa)
FT                   /transl_except= (pos:217. .219,aa:Xaa)
FT                   /transl_except= (pos:253. .255,aa:Xaa)
FT                   /transl_except= (pos:265. .270,aa:Xaa-Xaa)
FT                   /transl_except= (pos:394. .396,aa:Xaa)
FT                   /transl_except= (pos:547. .549,aa:Xaa)
FT                   /transl_except= (pos:565. .567,aa:Xaa)
FT                   /transl_except= (pos:616. .618,aa:Xaa)
FT                   /transl_except= (pos:649. .651,aa:Xaa)
FT                   /transl_except= (pos:682. .684,aa:Xaa)
FT                   /transl_except= (pos:709. .711,aa:Xaa)
FT                   /transl_except= (pos:790. .792,aa:Xaa)
FT                   /transl_except= (pos:841. .843,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:880. .882,aa:Xaa)
FT                   /transl_except= (pos:907. .909,aa:Xaa)
FT                   /transl_except= (pos:952. .954,aa:Xaa)
FT                   /transl_except= (pos:982. .984,aa:Xaa)
FT                   /transl_except= (pos:988. .990,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
```

```
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98739.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Claim 5; Page 170-172; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the cDNA sequence for human apo-dystrophin-4
CC   containing a plurality of stop codons some of which may be read through
CC   due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 996;  DB 6;  Length 996;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  996;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996


RESULT 2
ABK86462
ID   ABK86462 standard; cDNA; 1230 BP.
XX
AC   ABK86462;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human Apo-dystrophin-4 full length cDNA.
XX
KW   Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
KW   protein truncation; muscular dystrophy; leukaemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1230
FT                   /*tag=  a
FT                   /product= "Apo-dystrophin-4"
FT                   /partial
FT                   /note= "No start or stop codon shown"
FT                   /transl_except= (pos:142. .144,aa:Xaa)
FT                   /transl_except= (pos:241. .244,aa:Xaa)
FT                   /transl_except= (pos:282. .284,aa:Xaa)
FT                   /transl_except= (pos:421. .423,aa:Xaa)
FT                   /transl_except= (pos:436. .438,aa:Xaa)
```

```
FT                   /transl_except= (pos:451. .453,aa:Xaa)
FT                   /transl_except= (pos:487. .489,aa:Xaa)
FT                   /transl_except= (pos:499. .504,aa:Xaa-Xaa)
FT                   /transl_except= (pos:628. .630,aa:Xaa)
FT                   /transl_except= (pos:781. .783,aa:Xaa)
FT                   /transl_except= (pos:799. .801,aa:Xaa)
FT                   /transl_except= (pos:850. .852,aa:Xaa)
FT                   /transl_except= (pos:883. .885,aa:Xaa)
FT                   /transl_except= (pos:916. .920,aa:Xaa)
FT                   /transl_except= (pos:945. .947,aa:Xaa)
FT                   /transl_except= (pos:1024. .1026,aa:Xaa)
FT                   /transl_except= (pos:1075. .1077,aa:Xaa)
FT                   /transl_except= (pos:1084. .1086,aa:Xaa)
FT                   /transl_except= (pos:1114. .1116,aa:Xaa)
FT                   /transl_except= (pos:1141. .1143,aa:Xaa)
FT                   /transl_except= (pos:1186. .1188,aa:Xaa)
FT                   /transl_except= (pos:1216. .1218,aa:Xaa)
FT                   /transl_except= (pos:1222. .1224,aa:Xaa)
FT                   /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN   GB2368064-A.
XX
PD   24-APR-2002.
XX
PF   16-JAN-2001; 2001GB-00001124.
XX
PR   30-SEP-2000; 2000US-0237079P.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA   (BARB/) BARBER E.
XX
PI   Barber E;
XX
DR   WPI; 2002-429042/46.
DR   P-PSDB; AAU98729.
XX
PT   New human regulatory polynucleotide, useful for treating disorders
PT   associated with protein truncation, particularly muscular dystrophy, and
PT   related peptides and antibodies.
XX
PS   Disclosure; Fig 4; 222pp; English.
XX
CC   The invention relates to a polynucleotide (I) comprising, or consisting
CC   of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC   functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC   as ABK86497). Also included are polynucleotides that hybridise to either
CC   strand of (I), a vector containing (I), a cell containing (I) or the
CC   vector, proteins and peptides encoded by (I), a protein homologous with
CC   human dystrophin that is expressed on cell surfaces in vivo antibodies
CC   (Ab) specific for the protein and method of screening for leukemia cells
CC   by analysing DNA for presence of (I) or by detecting presence of (II).
CC   The apo-dystrophin-4 inversion sequence is a regulatory element that
CC   controls expression (transcription and translation) of associated DNA,
CC   and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC   sequence is used in gene therapy of diseases associated with truncation
CC   of proteins, particularly muscular dystrophy and also leukaemia, but more
CC   generally (I) is a regulatory sequence used to control expression of any
CC   attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC   encoded by (I), can be used to screen for leukaemic cells and related
CC   diseases. Antibodies raised against (II) can be used therapeutically, to
CC   inhibit (II) activity, also to detect (II) in screening assays. The
CC   present sequence is the full length cDNA sequence for human apo-
CC   dystrophin-4 containing a plurality of stop codons some of which may be
CC   read through due to the presence of (I) in the apo-dystrophin-4 gene
XX
SQ   Sequence 1230 BP; 404 A; 189 C; 259 G; 378 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 996;  DB 6;  Length 1230;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  996;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        235 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 294

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        295 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 354

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        355 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 414

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        415 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 474

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        475 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 534

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        535 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 594

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        595 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 654

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        655 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 714

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        715 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 774

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        775 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 834

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        835 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTTGTGAAGGGTAGTG 894

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        895 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 954

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        955 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 1014

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1015 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 1074

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1075 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 1134

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1135 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 1194

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db       1195 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 1230

RESULT 3
ABK86463
ID   ABK86463 standard; cDNA; 1234 BP.
XX
AC   ABK86463;
XX
```

DT 27-AUG-2002 (first entry)
XX
DE Human Apo-dystrophin-4 full cDNA fragment #1.
XX
KW Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;
KW protein truncation; muscular dystrophy; leukaemia.
XX
OS Homo sapiens.
XX
PN GB2368064-A.
XX
PD 24-APR-2002.
XX
PF 16-JAN-2001; 2001GB-00001124.
XX
PR 30-SEP-2000; 2000US-0237079P.
XX
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA (BARB/) BARBER E.
XX
PI Barber E;
XX
DR WPI; 2002-429042/46.
XX
PT New human regulatory polynucleotide, useful for treating disorders
PT associated with protein truncation, particularly muscular dystrophy, and
PT related peptides and antibodies.
XX
PS Disclosure; Fig 9; 222pp; English.
XX
CC The invention relates to a polynucleotide (I) comprising, or consisting
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
CC as ABK86497). Also included are polynucleotides that hybridise to either
CC strand of (I), a vector containing (I), a cell containing (I) or the
CC vector, proteins and peptides encoded by (I), a protein homologous with
CC human dystrophin that is expressed on cell surfaces in vivo antibodies
CC (Ab) specific for the protein and method of screening for leukemia cells
CC by analysing DNA for presence of (I) or by detecting presence of (II).
CC The apo-dystrophin-4 inversion sequence is a regulatory element that
CC controls expression (transcription and translation) of associated DNA,
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
CC sequence is used in gene therapy of diseases associated with truncation
CC of proteins, particularly muscular dystrophy and also leukaemia, but more
CC generally (I) is a regulatory sequence used to control expression of any
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
CC encoded by (I), can be used to screen for leukaemic cells and related
CC diseases. Antibodies raised against (II) can be used therapeutically, to
CC inhibit (II) activity, also to detect (II) in screening assays. The
CC present sequence is a human apo-dystrophin-4 fragment matching a region
CC of mouse genomic DNA in the Apo-dystrophin-4 gene region
XX
SQ Sequence 1234 BP; 405 A; 190 C; 260 G; 379 T; 0 U; 0 Other;

Query Match 100.0%; Score 996; DB 6; Length 1234;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT 298

Qy         61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 358

Qy        121 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT 418

Qy        181 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA 478

Qy        241 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA 538

Qy        301 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG 598

Qy        361 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 658

Qy        421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 718

Qy        481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 778

Qy        541 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 838

Qy        601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 898

Qy        661 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 958

Qy        721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG 1018

Qy        781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 1078

Qy        841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 1138

Qy        901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA 1198

Qy        961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
              ||||||||||||||||||||||||||||||||||||
Db       1199 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 1234
```

RESULT 4
ABT33375
ID ABT33375 standard; DNA; 2463 BP.
XX
AC ABT33375;
XX
DT 22-MAY-2003 (first entry)
XX
DE NOVX DNA sequence SEQ ID No 64.
XX
KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;

KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
KW immunogen; non-human transgenic animal; gene therapy; gene; ds.
XX
OS Unidentified.
XX
PN WO200281517-A2.
XX
PD 17-OCT-2002.
XX
PF 22-JAN-2002; 2002WO-US002064.
XX
PR 19-JAN-2001; 2001US-0262892P.
PR 23-JAN-2001; 2001US-0263598P.
PR 24-JAN-2001; 2001US-0263799P.
PR 25-JAN-2001; 2001US-0264117P.
PR 25-JAN-2001; 2001US-0264139P.
PR 26-JAN-2001; 2001US-0264478P.
PR 30-JAN-2001; 2001US-0263351P.
PR 02-MAR-2001; 2001US-0272870P.
PR 14-MAR-2001; 2001US-0275927P.
PR 14-MAR-2001; 2001US-0275990P.
PR 15-MAR-2001; 2001US-0276449P.
PR 20-MAR-2001; 2001US-0277358P.
PR 23-MAR-2001; 2001US-0278151P.
PR 29-MAR-2001; 2001US-0279857P.
PR 20-APR-2001; 2001US-0285140P.
PR 20-APR-2001; 2001US-0285141P.
PR 30-APR-2001; 2001US-0287484P.
PR 17-MAY-2001; 2001US-0291701P.
PR 08-JUN-2001; 2001US-0296960P.
PR 10-JUL-2001; 2001US-0304353P.
PR 10-JUL-2001; 2001US-0304355P.
PR 12-JUL-2001; 2001US-0304886P.
PR 09-AUG-2001; 2001US-0311289P.
PR 13-AUG-2001; 2001US-0311975P.
PR 16-AUG-2001; 2001US-0312937P.
PR 18-OCT-2001; 2001US-0330227P.
PR 29-NOV-2001; 2001US-0334198P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;
XX
DR WPI; 2003-058504/05.
DR P-PSDB; ABJ37909.
XX
PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.
XX
PS Claim 9; Page 171; 672pp; English.
XX
CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
CC acids can be used in gene therapy. This polynucleotide sequence
CC represents a NOVX DNA sequence of the invention
XX
SQ Sequence 2463 BP; 691 A; 599 C; 580 G; 593 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 7; Length 2463;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2009 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 2068

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2069 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 2128

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2129 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 2188

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2189 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 2248

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2249 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 2308

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2309 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 2368

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2369 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 2418
```

RESULT 5
ABK81998
ID ABK81998 standard; DNA; 5339 BP.
XX
AC ABK81998;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR2-R21.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.

OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Example 6; Fig 13; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5339 BP; 1638 A; 1191 C; 1187 G; 1323 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 5339;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4551 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4610

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4611 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4730

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4731 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4790

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4791 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4850

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4851 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 4900
```

RESULT 6
ABK81997
ID ABK81997 standard; DNA; 5417 BP.
XX
AC ABK81997;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR4-R23.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 12; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5417 BP; 1700 A; 1192 C; 1182 G; 1343 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 5417;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4628

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4629 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4688

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4689 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4748

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4749 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4808

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4809 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4868

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4869 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4928

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 4978
```

RESULT 7
ABK81999
ID ABK81999 standard; DNA; 5462 BP.
XX
AC ABK81999;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR2-R21+H3.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 14; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5462 BP; 1668 A; 1225 C; 1212 G; 1357 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 5462;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 4673

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 4733

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4734 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4793

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4794 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 4853

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4854 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 4913

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4914 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 4973

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4974 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 5023
```

RESULT 8
ABK82000
ID ABK82000 standard; DNA; 8689 BP.
XX
AC ABK82000;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaH2-R19.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 15; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 8689;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 6044

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 6104
```

```
Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6105 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 6164

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6165 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 6224

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6225 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 6284

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6285 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 6344

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       6345 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 6394
```

RESULT 9
ABK82002
ID ABK82002 standard; DNA; 11443 BP.
XX
AC ABK82002;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR9-R16.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 17; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 11443;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8739 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 8798

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 8858

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8859 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 8918

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8919 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 8978

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8979 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 9038

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9039 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 9098

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       9099 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 9148
```

RESULT 10
AAN90338
ID AAN90338 standard; cDNA; 12923 BP.
XX
AC AAN90338;
XX
DT 29-MAR-1992 (first entry)
XX
DE Sequence of human muscular dystrophy (MD) cDNA.
XX
KW Dystrophin; muscular dystrophy; probe; antibody; diagnosis; prenatal;
KW heterozygote; gene therapy; genetic screening; foetal screening; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 209. .12923
FT /*tag= a
XX
PN WO8906286-A.
XX
PD 13-JUL-1989.
XX
PF 16-DEC-1988; 88WO-US004504.
XX
PR 22-DEC-1987; 87US-00136618.
XX
PA (CHIL-) CHILDRENS MED CENT.
XX
PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;
XX
DR WPI; 1989-220587/30.
DR P-PSDB; AAP90373.
XX
PT Muscular dystrophy gene - used for prepn. of probes, dystrophic
PT polypeptide and antibodies for diagnosis and therapy of muscular
PT dystrophy.
XX
PS Disclosure; Fig 5; 68pp; English.
XX
CC The inventors claim an MD probe comprising a purified ss NA SQ which
CC hybridises to at least a part of the MD gene; pure dystrophin (DS)

```
CC   polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
CC   probes are equal to or greater than 10b of one of 12 cDNA sequences
CC   deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd gene
XX
SQ   Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T; 0 U; 7 Other;

  Query Match             41.2%;  Score 410;  DB 1;  Length 12923;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-190;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662


RESULT 11
ABK81959
ID   ABK81959 standard; DNA; 13957 BP.
XX
AC   ABK81959;
XX
DT   13-AUG-2002  (first entry)
XX
DE   cDNA encoding human dystrophin.
XX
KW   Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW   Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200229056-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-US031126.
XX
PR   06-OCT-2000; 2000US-0238848P.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Chamberlain JS,  Harper SQ;
XX
DR   WPI; 2002-435334/46.
XX
PT   A composition for preparing therapeutic drugs, has a mini-dystrophin
PT   peptide comprising a specific number of spectrin-like repeat domains, or
PT   a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS   Example 2; Fig 1; 145pp; English.
XX
CC   The invention describes a composition comprising a mini-dystrophin
CC   peptide comprising a spectrin-like repeat domain, where the domain
CC   comprises n spectrin-like repeats, and contains no more than n spectrin-
CC   like repeats, where n is an even number between 4-24, or a nucleic acid
CC   encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC   polynucleotide encoding it is useful as a medicament, for preparing a
CC   drug for therapeutic application and in the preparation of a composition
CC   for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC   (DMD). This sequence represents a human dystrophin polynucleotide
CC   sequence used in the creation of the mini-dystrophin peptides of the
CC   invention
XX
SQ   Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

  Query Match             41.2%;  Score 410;  DB 6;  Length 13957;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-190;
  Matches  410;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662


RESULT 12
ABT10904
ID   ABT10904 standard; cDNA; 13957 BP.
XX
AC   ABT10904;
XX
DT   04-DEC-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 1038.
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
```

PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR WPI; 2002-674803/72.
XX
PT Diagnosing breast cancer in a patient comprises detecting the level of
PT gene expression in cell or tissue samples, where a differential gene
PT expression is indicative of breast cancer.
XX
PS Claim 1; SEQ ID NO 1038; 260pp + Sequence Listing; English.
XX
CC The present invention relates to methods of diagnosing breast cancer in a
CC patient, which comprise detecting the level of expression in a tissue
CC sample of two or more genes selected from those shown in ABT09867-
CC ABT11112, where a differential expression of the genes indicates breast
CC cancer. The methods are useful in diagnosing, treating, detecting the
CC progression, and in monitoring treatment of breast cancer in patients.
CC The methods are also useful as a screening tool for agents that modulate
CC the onset or progression of breast cancer. The breast cancer genes may be
CC used as diagnostic markers for the prediction or identification of the
CC malignant state of breast tissue, for confirming the type and progression
CC of cancer, and for drug screening and assays. The present sequence is a
CC coding sequence of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub.published_pct_sequences
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662
```

RESULT 13
ABN95786
ID ABN95786 standard; DNA; 13957 BP.
XX
AC ABN95786;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2284 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 2284; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy       510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy       570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy       630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy       690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy       750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy       810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
```

```
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662
```

RESULT 14
ABS69900
ID ABS69900 standard; DNA; 13957 BP.
XX
AC ABS69900;
XX
DT 21-NOV-2002 (first entry)
XX
DE Human dystrophin gene.
XX
KW Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;
KW adenosine deaminase deficiency; severe combined immune deficiency; PAH;
KW beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;
KW low density lipoprotein gene; familial hypercholesterolaemia;
KW hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;
KW phenylalanine hydroxylase gene; gene therapy; phenylketonuria;
KW dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;
KW human cystic fibrosis transmembrane conductance regulator gene;
KW antianaemic; antilipaemic; nootropic; cytostatic; dermatological; human;
KW alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;
KW lysosomal arylsulfatase A; omithine transcarbamylase; ARSA; OTC; NP;
KW purin nucleoside phosphorylase; gene; ds.
XX
OS Homo sapiens.
XX
PN US2002102731-A1.
XX
PD 01-AUG-2002.
XX
PF 12-FEB-2001; 2001US-00782378.
XX
PR 02-OCT-2000; 2000US-0237747P.
XX
PA (UYNY) UNIV NEW YORK STATE RES FOUND.
XX
PI Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;
XX
DR WPI; 2002-690619/74.
XX
PT Producing vector, by introducing vector having nucleotide sequence,
PT adenovirus inverted terminal repeats and packaging sequence, and adeno-
PT associated virus terminal repeat, into cell, and culturing cell.
XX
PS Disclosure; Page 122-128; 191pp; English.
XX
CC The present invention relates to a new method of producing a vector. The
CC method involves introducing recombinant vector having nucleotide sequence
CC (NS) having 5' and 3' end, left and right inverted terminal repeats of
CC adenovirus flanking NS, adenovirus packaging sequence linked to inverted
CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'
CC end of NS, into cell expressing adenovirus early gene lacking from vector
CC ; and culturing cell to produce another vector. The method is useful for
CC generating vectors, especially mAd vectors. The method is useful in
CC transferring nucleotide sequences of interest into a cell, for gene
CC transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.
CC The nucleotide sequences are useful for treating diseases associated with
CC it, i.e. adenosine deaminase gene associated with adenosine deaminase
CC deficiency with severe combined immune deficiency, beta-chain of
CC haemoglobin gene associated with beta-thalassaemia and sickle cell
CC disease, receptor for low density lipoprotein gene associated with
CC familial hypercholesterolaemia, hypoxanthine-guanine
CC phosphoribosyltransferase associated with Lesch-Nyhan syndrome,
CC phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,
CC dystrophin gene associated with muscular dystrophy, and human cystic
CC fibrosis transmembrane conductance regulator gene associated with cystic
CC fibrosis. The present nucleic acid sequence represents a human disease
CC gene sequence that was used in the methods of the invention
XX
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTC 11312

Qy        510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATT 11372

Qy        570 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11373 CCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

Qy        630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11433 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCT 11492

Qy        690 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11493 AAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAA 11552

Qy        750 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11553 AAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATA 11612

Qy        810 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11613 TGGAACGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 11662
```

RESULT 15
ABK81996
ID ABK81996 standard; DNA; 2691 BP.
XX
AC ABK81996;
XX
DT 13-AUG-2002 (first entry)
XX
DE Human dystrophin 3' untranslated region.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; human; ds.
XX
OS Homo sapiens.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238848P.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 10; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid

CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a human dystrophin polynucleotide
CC sequence used in the creation of the mini-dystrophin peptides of the
CC invention
XX
SQ Sequence 2691 BP; 860 A; 448 C; 461 G; 922 T; 0 U; 0 Other;

Query Match 39.8%; Score 396; DB 6; Length 2691;
Best Local Similarity 100.0%; Pred. No. 1.1e-183;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        464 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCA 60

Qy        524 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTT 120

Qy        584 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATAT 180

Qy        644 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 240

Qy        704 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAA 300

Qy        764 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTG 360

Qy        824 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 859
              |||||||||||||||||||||||||||||||||||
Db        361 GGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGA 396
```

Search completed: April 6, 2004, 15:10:53
Job time : 474.067 secs